Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US04/020805

International filing date: 28 June 2004 (28.06.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/483,463
Filing date: 27 June 2003 (27.06.2003)

Date of receipt at the International Bureau: 27 August 2004 (27.08.2004)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1211887

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 16, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/483,463*
FILING DATE: *June 27, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/20805*

Certified by

Jon W Dudas

Acting Under Secretary of Commerce
for Intellectual Property
and Acting Director of the U.S.
Patent and Trademark Office

175 U.S. PTO
06/27/03

Please type a plus sign (+) inside this box → [+]

PTO/SB/16 (5-03)
Approved for use through 4/30/2003. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

17507 U.S. PTO
60/483403
06/27/03

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
John S.	McGeachie	21 Bishops Way, North Reading, MA 01864 USA

☐ Additional inventors are being named on the ____ separately numbered sheets attached hereto

TITLE OF THE INVENTION (280 characters max)

System and Method for Establishing Connection-Based Personal Link Lists

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☒ Customer Number 23630 → *Place Customer Number Bar Code Label here*

OR *Type Customer Number here*

☒ Firm or Individual Name	McDermott, Will & Emery				
Address	28 State Street				
Address					
City	Boston	State	MA	ZIP	02109
Country	USA	Telephone	617-535-4037	Fax	617-535-3800

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒ Specification *Number of Pages* 46
☒ Drawing(s) *Number of Sheets* 7
☐ Application Data Sheet. See 37 CFR 1.76
☐ CD(s), Number
☒ Other (specify) Express Mail Cover Sheet and Ackn. Postcard

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT *(check one)*

☒ Applicant claims small entity status. See 37 CFR 1.27.
☐ A check or money order is enclosed to cover the filing fees
☒ The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number 50-1133
☐ Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($) $80.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.
☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE

TYPED or PRINTED NAME David M. Mello

TELEPHONE 617-535-4037

Date 06/27/03

REGISTRATION NO. *(if appropriate)* 43,799

Docket Number: 58169-025

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the complete provisional application to the PTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Mail Stop Provisional Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need asssitance in completing the form, call 1-800-PTO-9199 and select option 2.

CERTIFICATE OF MAILING BY "EXPRESS MAIL" (37 CFR 1.10) Applicant(s): John S. McGeachie			Docket No. 58169-025 (ORBK-002PR)
Serial No. TBA	Filing Date June 27, 2003	Examiner TBA	Group Art Unit TBA

Invention: **SYSTEM AND METHOD FOR ESTABLISHING CONNECTION-BASED PERSONAL LINK LISTS**

I hereby certify that the following correspondence:

New Provisional Patent Application, Addendum to Patent Application, Drawings (7 sheets), Provisional Application For Patent Cover Sheet (Small Entity), and Acknowledgment Postcard.

(Identify type of correspondence)

is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on

June 27, 2003
(Date)

Weiguo Chen
(Typed or Printed Name of Person Mailing Correspondence)

Weiguo Chen
(Signature of Person Mailing Correspondence)

EV332174757US
("Express Mail" Mailing Label Number)

Note: Each paper must have its own certificate of mailing.

P06A/REV02

System and Method for Establishing Connection-Based Personal Link Lists

Cross-Referenced Applications

This application claims the benefit U.S. application Serial No. 60/211,725, filed June 15, 2000, the disclosure of which is hereby incorporated by reference into this application.

Field of the Invention

The present invention relates generally to a method of and system for determining connections between parties and, more particularly, to a connection searching method and system in which a user is capable of entering a source party and a target party and searching a host database to obtain lists of people or entities through which the source and target parties are connected. The system also is capable of determining a number of connections that are associated with one party.

Background of the Invention

It is well known that personal contacts are advantageous when conducting transactions between parties. However, determining the contacts of one party of a transaction the contacts of the other party of the transaction and what contacts those contacts have in common can be very difficult and time consuming. Currently, there is no efficient method or system for determining such contacts between parties of a transaction.

Summary of the Invention

The present invention is directed to a method of and system for determining connections between people which is efficient and effective. The system includes a host database which includes records of parties, including identification information, which is available from non-restricted sources. The identification information is arranged in a series of searchable data fields. A user connects to a website associated with the system and inputs a source party and a target party, for the purpose of finding a number of connections between the parties. The parties may be people or entities, such as companies, organizations, etc. The system searches the database for intermediate party records having at least one data field which includes identification information which is common to the identification information in at least one of the data fields of the source party record. The located party records are compared to the target party record to determine if any of the identification information in the intermediate party record is common to any of the identification information in the target party record. If there is a commonality, a list of the source party, intermediate party and target party is generated, including the records for each party, to show the connection path between the source party and the target party. If there are no commonalities between the intermediate party and the target party, further intermediate parties are located which have commonalities with the first intermediate party. The located party records are then compared to the target party record to determine if any of the identification information in the further intermediate part records are common to any of the identification information in the target party record. If there is a commonality, a list of the source party, intermediate parties and target party is generated, including the records for each party, to show the connection path between the source party and the target party. This process is repeated until no further connections are found or until a preset limit of connections is reached.

According to one aspect of the invention, a method of determining a connection between a source party and a target party includes:

A. constructing a host database, the host database including a plurality of records, each record including a number of data fields, each of the data fields

including identification information of a party, the identification information being extracted from non-restricted sources;

B. constructing a client database, the client database including a plurality of records, each record including a number of data fields, each of the data fields including identification information of a party, the identification information being extracted from a client's private sources;

C. receiving identification information of a source party and a target party;

D. identifying a record in the client database including identification information of the source party;

E. identifying a record in the host database including identification information of the target party;

F. searching the data fields in the records of at least one of the client database and the host database to locate identification information commonalities between the source party record and at least one intermediate party record;

G. searching the data fields in the records of at least one of the client database and the host database to locate identification information commonalities between the at least one intermediate party record and the target party record; and

H. upon locating at least one identification information commonality between the at least one intermediate party record and the target party record, generating a list including the at least one intermediate party record.

Step G may further include searching the data fields in the records of at least one of the client database and the host database to locate identification information commonalities between the at least one intermediate party records and further intermediate party records; and searching the data fields in the records of at least one of the client database and the host database to locate identification information commonalities between the further intermediate party records and the target party record.

The source party and the target party may be one of a person and an entity. The identification information may include personal and affiliation information of the party.

The identification information may include at least one of a person's name, the person's dates of employment with a company, the person's title within the company, the person's company name, the person's company address, the person's company SIC code, and the person's company ticker symbol. The identification information may include at least one of a company name, the company's address, the company's SIC code and the company's ticker symbol. The records stored on the client database may be a subset of the records stored on the host database.

According to another aspect of the invention, a method of determining a connection between a source party and a target party includes:

A. constructing a host database, the host database including a plurality of records, each record including a number of data fields, each of the data fields including identification information of a party, the identification information being extracted from non-restricted sources;

B. receiving identification information of a source party and a target party;

C. identifying a record in the host database including identification information of the source party;

D. identifying a record in the host database including identification information of the target party;

E. searching the data fields in the records of the host database to locate identification information commonalities between the source party record and at least one intermediate party record;

F. searching the data fields in the records of the host database to locate identification information commonalities between the at least one intermediate party record and the target party record; and

G. upon locating a data field commonality between the at least one intermediate party record and the target party record, generating a list of the at least one intermediate party record.

According to another aspect of the invention, a system for determining a connection between a source party and a target party includes a host system having a computer processor and associated memory. The host system includes a host database including a plurality of records, each record including a number of data fields, each of the data fields including identification information of a party, the identification information being extracted from non-restricted sources. The system also includes a client system having a computer processor and associated memory, the client system including a client database including a plurality of records, each record including a number of data fields, each of the fields including identification information of a party, the identification information being extracted from a client's private sources. The client system establishes a connection to the host system over the communication network and inputs identification information of a source party and a target party. The host system identifies a record in at least on of the client database and the host database including identification information of the source party and identifying a record in at least one of the client database and the host database including identification information of the target party; and the host system searching the data fields in the records to locate identification information commonalities between the source party record and at least one intermediate party record and searching the data fields in the records to locate identification information commonalities between the at least one intermediate party record and the target party record. Upon locating a identification information commonality between the at least one intermediate party record and the target party record, the host system generating a list of the at least one intermediate party record.

According to yet another aspect of the invention, a system for determining a connection between a source party and a target party includes a host system including a computer processor and associated memory and a user system including a computer processor and associated memory. The host system includes a database having a plurality of records, each record including a number of data fields, each of the data fields including identification information of a party, the identification information being extracted from non-restricted sources.

The user system is adapted for establishing a connection to the host system over a communication network and inputting identification information of a source party and a target party to the host system.

The host system identifies records in the database including identification information of the source party identification information of the target party and searches the data fields in the records to locate identification information commonalities between the source party record and at least one intermediate party record and searching the data fields in the records to locate identification information commonalities between the at least one intermediate party record and the target party record. Upon locating a identification information commonality between the at least one intermediate party record and the target party record, the host system generating a list of the at least one intermediate party record.

According to yet another aspect of the invention, a method of determining a connection between a source party and a target party includes:

A. receiving identification information of a source party and a target party;

B. identifying a record in the client database including identification information of the source party, the client database including a plurality of records, each record including a number of data fields, each of the data fields including identification information of a party;

C. identifying a record in the host database including identification information of the target party, the host database including a plurality of records, each record including a number of data fields, each of the data fields including identification information of a party;

D. searching the data fields in the records of at least one of the client database and the host database to locate identification information commonalities between the source party record and at least one intermediate party record;

E. searching the data fields in the records of at least one of the client database and the host database to locate identification information

commonalities between the at least one intermediate party record and the target party record; and

F. upon locating at least one identification information commonality between the at least one intermediate party record and the target party record, generating a list including the at least one intermediate party record.

According to yet another aspect of the invention, a method of determining a connection between a source party and a target party includes:

A. receiving identification information of a source party and a target party;

B. identifying a record in the host database including identification information of the source party, the host database including a plurality of records, each record including a number of data fields, each of the data fields including identification information of a party, the identification information being extracted from non-restricted sources;

C. identifying a record in the host database including identification information of the target party;

D. searching the data fields in the records of the host database to locate identification information commonalities between the source party record and at least one intermediate party record;

E. searching the data fields in the records of the host database to locate identification information commonalities between the at least one intermediate party record and the target party record; and

F. upon locating a data field commonality between the at least one intermediate party record and the target party record, generating a list of the at least one intermediate party record.

According to yet another aspect of the invention, a system for determining a connection between a source party and a target party includes a host system including a computer processor and associated memory and a user system including a computer processor and associated memory. The host system includes a database having a plurality of records, each record including a number of data fields, each of the data fields including identification information of a party. The user system is adapted for establishing a connection to the host

system over a communication network, the user system inputting identification information of a source party and a target party to the host system. The host system identifies records in the database including identification information of the source party identification information of the target party and searches the data fields in the records to locate identification information commonalities between the source party record and at least one intermediate party record and searching the data fields in the records to locate identification information commonalities between the at least one intermediate party record and the target party record. Upon locating a identification information commonality between the at least one intermediate party record and the target party record, the host system generating a list of the at least one intermediate party record.

Brief Description of the Drawings

The foregoing and other objects of this invention, the various features thereof, as well as the invention itself may be more fully understood from the following description when read together with the accompanying drawings in which:

Fig. 1 is a schematic diagram of a system for determining connections between parties in accordance with the present invention;

Fig. 2 is a flow diagram showing one embodiment of a method for determining connections between parties in accordance with the present invention;

Fig. is a flow diagram showing another embodiment of a method for determining connections between parties in accordance with the present invention;

Fig. 4 is a detailed schematic diagram of the system for determining connections between parties shown in Fig. 1;

Fig. 5 is a schematic diagram showing a list of connections determined according to the present invention;

Fig. 6 is a more detailed schematic diagram of records of the parties involved in one of the connections shown in Fig. 5; and

Fig. 7 is a schematic diagram showing identification information included in a record of an entity, according to the present invention.

Detailed Description

Fig. 1 shows a schematic diagram of a system 10 for determining connections between parties in accordance with a preferred embodiment of the present invention. The system 10 includes host system 12, user system 14 and client systems 16a-16c, all connected to a common communications network 18. While three client systems 16a-16c are shown in Fig. 1, it will be understood that as few as one client system may participate in the study, or many more than three may participate. Three client systems are shown in Fig. 1 for the purpose of example only. Preferably, the host system 12, user system 14 and client systems 16a-16c are each a personal computer such as an IBM PC or IBM PC compatible system or an APPLE MacINTOSH system or a more advanced database computer system such as an Alpha-based computer system available from Compaq Computer Corporation or SPARC Station computer system available from SUN Microsystems Corporation, although a main frame computer system can also be used. Preferably, the communications network 18 is a TCP/IP-based network such as the Internet or an intranet, although almost any well known LAN, WAN or VPN technology can be used.

In one preferred embodiment of the invention, the user system 14 is an IBM PC compatible system operating an operating system such as the Microsoft Windows™ operating system, and host system 12 is configured as a web server providing access to information such as web pages in HTML format via a protocol such as the HyperText Transport Protocol (http). The user system 14 and client systems 16a-16c include software to allow viewing of web pages, commonly referred to as a web browser, thus being capable of accessing web pages located on host system 12. Alternatively, user system 14 and client

system 16a-16c can be any wired or wireless device that can be connected to a communications network, such as an interactive television system, including WEBTV, a personal digital assistant (PDA) or a cellular telephone.

The method of and system for determining connections between parties will now be described with reference to Figs. 2-7. Fig. 4 is a schematic block diagram showing a more detailed diagram of the system 10 of Fig. 1. As shown in Fig. 4, host system 12 includes a host operation system and database 102 and a record matching engine 104. As shown in the figure, the client systems 16a-16c are separate entities, each having a firewall, represented by dashed line 124. The client systems 16a-16c are located on the client side 120 of the firewall 124 and the host system 12 is located on the host side 122 of the firewall 124. Each of client systems 16a-16c include a company database 110 in which contacts of employees and officers of the company are stored. Typical contact programs which are used for storing contacts include ACT®, Outlook®, Goldmine®, and Lotus Notes®. Each client system also includes an intermediate list 112, a client database 114 and a client interface 116. The operation of these components is described in detail below.

Fig. 2 is a flow diagram 20 which shows the method of determining connection between parties. In step 22, the host database is constructed. This involves populating the database with information about people and entities such as companies, organizations, etc. This information is extracted from non-restricted sources including the SEC database, Market Guide, IPO.com, company websites, news articles, press releases, etc. The information about each person or entity is arranged in a parsable record having a number of data fields. Identification information of the person or entity is input into an appropriate data field. The identification information input into the various data fields includes the name of the person or entity, the address of the person or entity, the person's position in the company, the person's dates of employment with companies the person has worked for, the ticker symbol of the company, the SIC code of the company, etc. Preferably, the majority of the information is obtained through an automated process, such a web crawler, that searches the

internet, extracts the appropriate data and inserts the data into the data fields to construct a record of the person or entity. Information not accessible to the automated process is input to the system manually. In this step, relationships between parties may be identified and links between related records established and saved in the database. This enables connections between parties to be included in the records of each party. This step may be repeated as often as necessary to keep the database updated.

In step 24, the client database 114 is constructed. First, the contact data included in the company database 110 is exported to the company list 112, and irrelevant contacts, such as personal contacts and non-business contacts, are eliminated. Redundant contacts are also eliminated. The company list 112 is input to record matching engine 104 where it is compared to the records included on host operation system and database 102. All contacts in the company list 112 that are also included in the host database 102 are stored in the same record form as the host database contacts and these records are saved in client database 114. This step may be repeated as often as necessary to keep the database updated. Accordingly, the data stored in the client database 116 is a subset of the data stored in host database 102. Known relationships between records in the client database 116 can be determined at this point and links between the related records implemented into the records. The information stored in the client database is proprietary to the client and is not accessible by outside parties. Contacts in the company list 112 which are not already on the host database 102 are not saved in the client database 114, since these contacts will not lead to further contacts on the host database 102.

Once the party records have been constructed and stored in the client database 114 and the host database 102, the process of determining connections between parties (people and/or entities) can be executed. In step 26, the host operation system 102 receives identification information of the source party and the target party, which typically are the names of the person or entity, from the client interface 116 of the client system 16 through a connection with the host system 12 via the internet 18. The record associated with source

party is then located in the client database 114 if it is stored there. If it is not, it is located in the host database 102, step 28. The record associated with the target party is also located in either the client database 114 or the host database 102. In step 30, the records in the client database 114 and host database 102 are searched by the host operation system to locate commonalities between the identification information in the data fields in the source party record and identification information in the data fields of the records stored in the databases. All intermediate party records which include commonalities with the source party record are identified as first stage intermediate party records. If relationship links between parties within the client database have been previously established, these links are used to locate the connections between the source party record and the first stage intermediate party record. The identification information in the data fields of the first stage intermediate party records are then compared to the identification information in the data fields of the target party record to locate first stage intermediate party records having commonalities with the target party record, step 32. If none of the first stage intermediate party records have any identification information commonalities with the target party record, step 34, the records in the databases are searched to locate further stage intermediate party records having identification information commonalities with the first stage intermediate party records, step 36. The identification information in the further stage intermediate party records is searched to determine if there are any commonalities between any of the data fields in the further stage intermediate party records and the target party record, step 32. Steps 32 through 36 are repeated until an intermediate party record is located which has identification information commonalities with the target party record. When this occurs, the host operation system 102 generates a list of the parties connecting the source party to the target party, step 38, and transmits the list to the client interface 116 via the internet 18. If a preset limit, which limits the number of unique connections found to a predetermined number, which may be set by the client when entering the source and target party information or by the host operation system, is met, step 40, the process ends. If the preset limit is not met, steps 32

through 36 are repeated until the preset limit number of unique connections is met.

An example connections list is schematically shown in Fig. 5. In this example, after the host database 102 and the client database 114 have been constructed, steps 22 and 24, Fig. 2, identification information for a source party 202 and a target party 204 are input to the host operation system 102 over the internet 18 through client interface 116. For simplicity, the entire record of each party is not shown in Fig. 5. Only the relevant identification information for the purposes of this example are shown. Upon receiving the identification information, which typically is the name of the people between whom a connection is to be determined, the records of the source party and the target party are identified in the client and/or host databases, step 28. In this example, the source party 202 is for J.F. who is the Chief Technology Officer of Company A. The target party record 204 is for L.S., the Chief Financial Officer of Company F. In step 30, the host database 102 is searched to locate intermediate party records having identification information commonalities with the source party record 202. In this case, the record 206 of C.O., which indicates that C.O. has identification information including a relationship with Company A as Chief Operating Officer is located. The remaining identification information of the record of C.O. is searched to determine whether there is a commonality between any of the identification information of C.O. and any of the identification information stored in the record of the target party, L.S., step 32. There is a commonality, since the record of C.O. indicates a relationship with Company F as Chief Technology Officer, step 34. Accordingly, a list including the source party record of J.F., the intermediate party record of C.O. and the target party record of L.S. is generated and sent to the client interface 116, step 38. In the list, all of the identification information included data included in the record of each party is available to the client.

A more detailed view of the source party record 202, the target party record 204 and the intermediate party record 206 is shown in Fig. 6. As shown in the figure, the records 202, 204 and 206 include data fields listing identification

information such as the name of the person, age, address and relationships to entities such as companies, association, etc. In the example, the commonality between the source party record 202 and the intermediate party record 206 found in step 30 is that both J.F. and C.O. have a relationship with Company A: J.F. as the present CTO and C.O. as a former COO. In step 32, the commonality between the intermediate party record 206 and the target party record 204 is located, namely the relationship of both parties with Company F. As shown in record 204, L.S. is the present CFO of Company F and C.O. is the present COO of Company F.

In Fig. 5, each entity with which the involved parties are associated is indicated by a dashed line. Connections between entities are referred to as hops. Since no entities other than the entities associated with the source party and the target party are needed to make the connection shown by double-dotted, dashed line 208, this connection is referred to as a "one-hop" connection. Other, multiple hop connections between the source party record 202 and the target party record 204 are shown in Fig. 5. Line 210 shows a "two hop" connection. Using the method described above, it is determined that the record of the source party J.F., 202 indicates a relationship between Company A and Company D based on the commonality that J.F. is associated with both companies. A further search in host database 102 indicates a relationship between the record 202 of J.F. and the record 212 of M.P., based on the commonality that both parties have a relationship with Company D. The record 212 of M.P. indicates a relationship with the target record 204 of L.S., based on the commonality that both parties have a relationship with Company F. Accordingly, since one intermediate entity (Company D) is included in the connection 210, this connection is referred to as a "two-hop" connection. Three-hop connections are shown by dotted line 220 and dotted dashed line 222. In this example, if the preset limit of connections is four, in step 40, Fig. 2, the process would end. It will be understood that the preset limit can be set to any number, although, in order to minimize processing time and cumbersome connection lists, the limit preferably is set to no more than 10.

While the example described above shows how connections between two people are generated, the system also determines connections between a person and an entity, such as a company or association; between an entity and a person; and between two entities. Upon constructing the client database 116, a record of the client entity is generated and stored in the client database 116. The host database 102, when being constructed, generates records of entities found in its search of the non-restricted sources in the same manner as the records for people described above. An example entity record 230 is shown in Fig. 7. As shown in the figure, entity record 230 comprises a number of data fields including identification information of the entity, including the entity name, ticker symbol, address and a list of its executives. When an entity is the source party and/or the target party, The same process shown in Fig. 2 is carried out, meaning that intermediate records, which may include records of people or entities, are located which include identification information which is common to the source and or target party records.

In an alternative embodiment, the host operation system and database 102 and the record matching engine 104 are replicated on the client database 114. In this embodiment, all of the operations described above are executed on the client system 16, thus allowing all execution to be local to the client system 16. Furthermore, the system 10 can be utilized to construct a list of connection that are associated with a single party. By inputting a single party to the host operation system and database 102, the searching function described above is executed and, in a first iteration, all records including identification information having commonalities with the source party are located and displayed. Depending on the scope of connections desired, numerous iterations of the search function can be executed in order to locate records of parties connected to the parties located in previous iterations.

While, as described above, the system 10 may be utilized by clients having a proprietary client database, it can also be utilized by a party which does not construct its own database. This process is shown in the flow diagram 240 of Fig. 3. In step 250, the user system 14, Figs. 1 and 4, establishes a connection

over the internet to the host system 12. The user system then enters the source party and the target party, step 252. The host operation system 102 identifies the records associated with the source party and the target party in the host database, step 254. Once the source party record and the target party record are found, steps 256 through 266 are executed, which are identical to steps 30 through 40 shown in flow diagram 20 of Fig. 2.

Accordingly, the present invention enables connections between people and entities to be determined using a convenient and efficient database construction and search tool. The invention is able to provide information about connections between parties based on commonalities in the identification information associated with each of the people and entities. The system can also be used simply for browsing through connections between parties and for obtaining the identification information associated with the record for a particular party. While the application has been described in connection with an example using businesses and business people as the parties, it will be understood that any party could utilize the connection-determining feature of the present invention and be the subject matter, including schools, civic groups, churches, organizations, associations, families, agencies, neighborhoods, etc., and the people who populate such groups.

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The present embodiments are therefore to be considered in respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of the equivalency of the claims are therefore intended to be embraced therein.

For example, the system may include the various features and capabilities described in the **Addendum** ("*PrivateLink*") attached hereto, which forms part of this description and is incorporated herein by reference.

ADDENDUM

PRIVATE LINK

With the foregoing system and method, features may be included that allow a user to generate a list of persons or entities (including groups of persons or groups of lists) that can function as a starting point for a connections query or request. This system and method can be referred to as PrivateLink™ (offered by Orion's Belt). And, a user's private list can be referred to as the user's the user's PrivateLink list.

When the connections server and DB (or connections system) receives a query including a PrivateLink list and an endpoint, the system generates information representing the connections to the endpoint for each member of the PrivateLink list, and returns this to the user.

In other forms, rather than a single endpoint, a list of endpoints could be used (i.e., an endpoint list). In such a case, the system generates connections between each member in the PrivateLink list and each member in the endpoint list, to the extent such connections exist.

In yet another form, a user may enter a single starting point and an endpoint list. In such a case the system generates connections from the starting point to each endpoint in the endpoint list, to the extent such connections exist.

The following text describes these features more fully.

The Orion's Belt platform has several proprietary components:

- The Orion's Belt Database – which contains information derived from public-domain sources about people and entities with which they are associated (current and past)
- PrivateLink – which provides a secure mechanism for clients to link their confidential contact information with the Orion's Belt Database, and
- The application – which includes software to extract and parse relevant content from a variety of sources, coupled with connection algorithms to search for and identify linkages between people and/or entities.

The System (aka "Orion's Belt")

Orion's Belt offers users specialized query capabilities against a proprietary database. Our product has three essential components:

- **Connection algorithms, which search** for and identify connections between people and/or entities, either within the Orion's Belt Database or between a client's contacts and the Orion's Belt Database.
- **ClientLink**, which provides a proprietary, secure mechanism for clients to link their own information with the Orion's Belt Database. These links can leverage the full spectrum of a client's contacts and are available only to users authorized by that client. ClientLink is presented to the user as *PrivateLink* to reinforce the confidentiality of this information.

A **proprietary database**, which contains information (derived from public-domain sources) about entities and the people associated with them. This database is accessible to all Orion's Belt customers.

Functionality

Connect is the central differentiating component of what Orion's Belt has to offer. The system allows users to specify both the desired endpoints of a connection – people, entities or ClientLink – and the degrees of separation. It may also provide for an enhanced graphical display, and allow filtering according to the presence of specific people or entities in the connection paths (e.g., only show links with Michael Jordan in the path).

Other optional features include functions to:

- develop metrics to rank connections according to their probable value,
- permit the user to assign a personal weighting factor to connections, and
- display connections in priority order.

ClientLink allows our customers to integrate knowledge about their own connections and networks of relationships with the Orion's Belt database. For a multi-user subscriber, ClientLink will entail utilization of sophisticated permission protocols regarding access to information by individual users. Clients can indicate the existing people and entities in the Orion's Belt database with which they have relationships. Additionally, the system

can enable clients to "fill in the blanks" with ClientLink, i.e., add additional information about relationships between people and entities. All of the ClientLink information is kept proprietary to the specific subscriber.

Browse displays first-order relationships for a specified person, entity or PrivateLink list. An optional feature, "*Explore*", allows the user to easily determine concentric, expanding relationships radiating out from a central *ending* point, whether a person or an entity. Extended Browse capabilities allow searching along a number of parameters such as functional position (e.g., CEO) or education (e.g., MIT alumni).

ClientLink Integration

Synchronizing each customer's PrivateLink data with Orion's Belt is the process whereby names in a client's contact list are matched to names in our database. Then, client subscribers can connect from their personal or corporate contacts to the decision-makers in Orion's Belt.

The system accommodates this synchronization through a variety of mechanisms, ranging from plug-ins for popular CRM and contact management systems to customized extraction.

2. ClientLink

Client Link is a proprietary feature which links a client's own contacts (e.g., customers, referral sources, vendors, etc.) with the OBI (or connections) database in order to produce the most effective links for each client. This feature allows a user to specify in advance the people or entities in the OBI database which are to be used as sources for a connection, thus eliminating the need to specify a unique starting point for each connection request.

Individual users' lists can be part of a group, and connections can be requested using groups as a starting point. This feature allows client users to request connections from their own or from their colleagues' contacts, depending on the flexibility of each client's protocols regarding access to lists. Client administrators have wide latitude in setting up groups, so that connections can be requested from an office, a region, a practice, or an entire organization. Security protocols prevent any client from accessing another client's ClientLink data.

OBI will customize ClientLink for each client during the installation process. This includes, for example, determining the most effective way to make existing contact lists (e.g., from common contact management or CRM products) accessible to the OBI system, identifying client protocols regarding users' lists, and working with the client administrator to establish the group/list structure. In the system, a user's ClientLink list is called a PrivateLink.

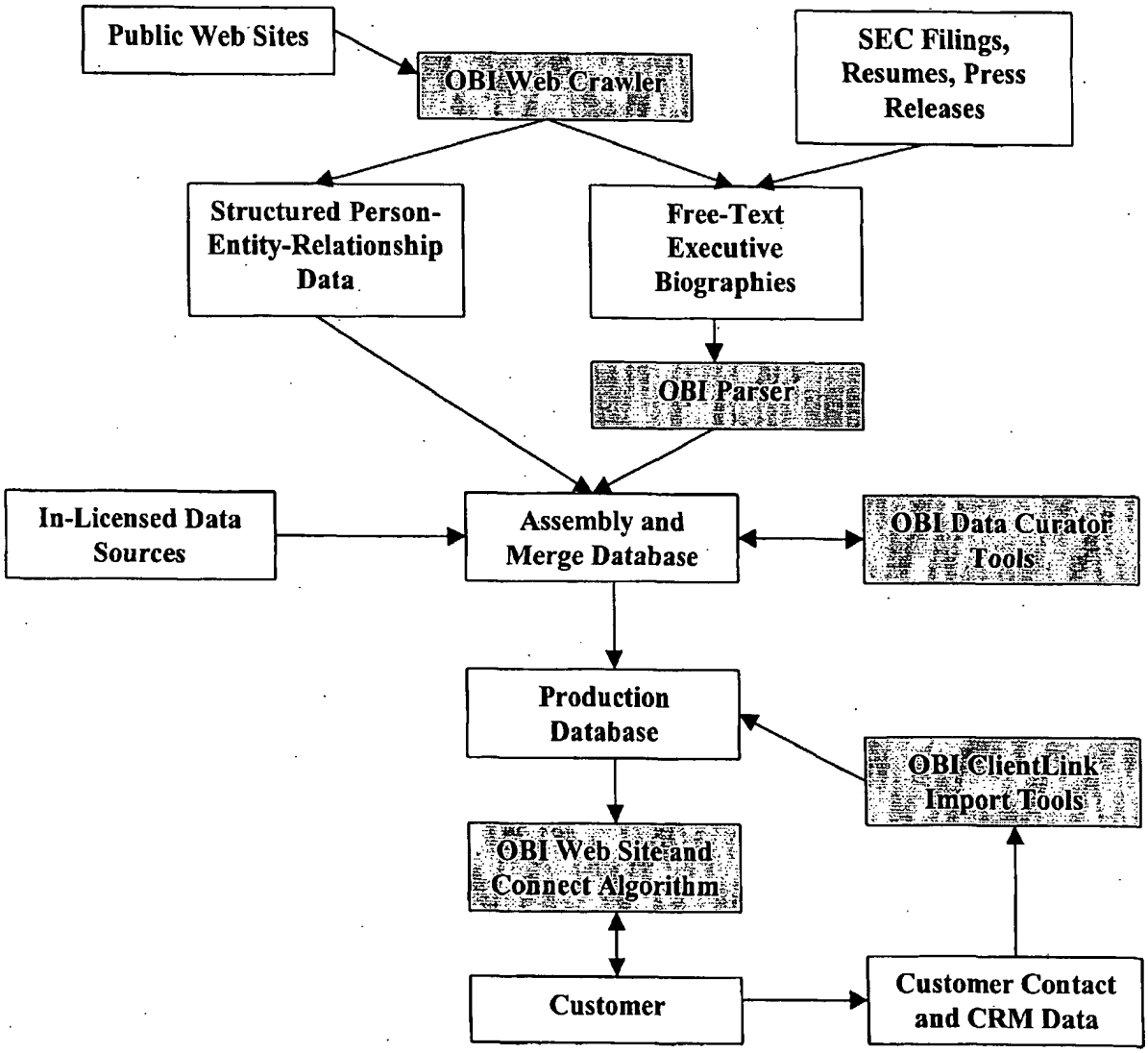
Users can populate their PrivateLink at the time of installation by extracting data from their current contact lists, or they can manually enter data into their PrivateLink as they use Orion's Belt (i.e., the connection system).

I. Technology

This technology contains two overall components:

- **Data-collection**—integrate data from multiple sources, verify, and load into the Orion's Belt database
- **Connection-finding**—search for links between people or entities and graphically display the results

Its components are linked in an overall information architecture, shown schematically below:



Data-Collection Technology

The Orion's Belt database contains information about entities, people, and the relationships among them:

Entities – companies and other organizations (e.g., "IBM")
People – individuals (e.g., "Louis V. Gerstner, Jr.")
Relationships – an affiliation and associated time period (e.g., "CEO from 1993")

This information is derived from publicly available sources (offered either free or by subscription) by a combination of automated methods with minimal manual intervention. The OBI database is populated via a four-step process:

1. A **web crawler** downloads information from public web sites or SEC filings, identifying information in headings and tables.
2. A proprietary OBI **parser** analyses the data and assembles information about entities, people, relationships, and dates.
3. A **data load** transfers this information into the OBI database.
4. Continuing **updates** keep the database current.

Web crawlers

Web crawlers are used to find and collect data about entities and the individuals associated with them. These data can be found at company web sites, SEC filings, and a variety of other sources, such as press releases. This data gathering process uses a combination of readily available tools (e.g., Wget) and ad-hoc OBI software. The Web crawler can identify some kinds of data relevant to OBI by its relationship to headings and tables on the HTML page.

Parser

For public corporations, the most useful sources of information—such as SEC filings or company web sites—generally contain "Executive Biographies", biographical paragraphs that provide background and supplementary data about each person associated with a particular corporation. These paragraphs are analyzed by a collection of computer programs we call the "parser" to identify entities, people, and relationships among them. An example of a paragraph from an SEC filing for the TALX Corporation is shown below:

> "Mr. Canfield has been President and Chief Executive Officer and a director of the Company since 1986 and has been Chairman of the Board of Directors since 1988. He had earlier become Chairman of the Board of EKI, Incorporated ("EKI"), which was acquired by the Company in fiscal 1994. For approximately 10 years, Mr. Canfield was President of Intech Group Inc. ("Intech Group"), until its acquisition by the Company in 1996, and from 1985 through 1989, Mr. Canfield was Chairman, and a principal shareholder of Noetic Technologies Corp., an engineering software company which was purchased by MacNeal-Schwendler Corporation in 1989. Prior to that, Mr. Canfield was one of two founders of Financial Data Systems, Inc. which was started in 1968. In 1980, the company, which provided services and turnkey systems to savings banks, was purchased by Citicorp. Mr. Canfield is a director of Jefferson Savings

Bancorp, Inc. Mr. Canfield holds a Bachelor of Science degree in Electrical Engineering from Purdue University and an M.B.A. degree from Washington University."

First, the parser partitions the paragraph into separate sentences. Then, the parser identifies entity names, people names, positions, and dates using a set of recognizer programs. Some of these elements are recognized heuristically (e.g., dates) while others are recognized by a combination of heuristics and by looking them up in a pre-defined list (e.g., entity names). The parser currently has a list of more than 64,000 entity names, entity name variants, and aliases (e.g., GE for General Electric Corporation).

Finally, the parser matches sentences containing recognized elements against a list of content patterns. If it finds a match, it uses the entity and position or title specified in the sentence to generate a corresponding relationship between an entity and a person. This relationship may also have start and end dates, if these were present in the sentence. If the parser cannot find a match between a sentence and its list of patterns, it creates a candidate pattern based on the sentence structure, but does not create a relationship. Instead, it records both sentence and candidate pattern to a log file for human review and, where appropriate, for manual input.

The parser can analyze about 90 sentences per second and takes about two hours to process all public companies listed on the NYSE, NASDAQ and AMEX exchanges. Currently, it accepts 30% to 40% of the information it encounters in free-text format. The acceptance rate will rise as we increase the number of content patterns, but it is unlikely to ever reach 100%; perhaps 60% to 75% is a realistic goal for well-written biographical paragraphs. The accuracy of the parsed data is very high—around 95%. Because of the high specificity of the parser, we are confident that it will be able to identify and extract correct relationships when they are mentioned in bodies of text where much of the content is on another topic (e.g., from press releases).

Some sources may be of such syntactic complexity or poor grammatical quality that the acceptance rate may be much lower. Even for well-written sources, however, eventually we will reach a point of diminishing returns, where the effort required to analyze sentences programmatically will exceed the effort required to do so manually. These remaining sentences will then be analyzed manually, using a small in-house editorial staff supplemented by an offshore data entry process.

From our experience to-date, it appears that larger, public companies tend to have better-written biographical paragraphs. These companies were our first priority for loading into the OBI database.

Data Load

When the parser has completed its work, the resulting output undergoes a modest amount of mostly automated follow-up processing to:

1. Identify and merge records to match up multiple references to a specific person from different sources by the same or closely related names

2. Identify and merge overlapping positions (different source paragraphs may refer to the same position with slightly different dates or with a different wording of the title)
3. Perform a sanity check on the parser output.

The final output is then loaded into an Oracle 8i relational database. As of 10/12/01, the Orion's Belt database contained approximately 26,000 entities, 75,000 people, and 250,000 associations between people and entities.

This database has been generated using computer software to extract information from electronically available data sources. Human input has been used to:

1. identify trustworthy information sources,
2. construct and correct errors in the parsing programs and their associated data tables, and
3. review the entity-name list (currently containing about 64,000 entities, and used by the parser to recognize valid entity names).

Update process

We keep the OBI database current in several ways:

1. Currently, our editorial staff makes corrections and data updates as we learn of them (e.g., from press releases, company web sites, etc).
2. The parsing technology can also compare current data (from our existing sources) against the database. If an entity/person/relationship set is in our database, but no longer in the source, an end date is inserted for that relationship. If entities, persons, or relationships are found in our sources, but not in the database, they are added to the database.
3. Also, statistical sampling can be used to verify the accuracy of the information loaded into our database.

We apply database updates daily. We allow database updates only from a single system with a secure connection to the database. We log all database changes (corrections, additions, and deletions) to an audit trail.

Connection-finding technology

The connection-related technology includes a user interface for access to the Orion's Belt database, and the algorithms required to find and to display connections between people and entities as requested by a user.

Access to Orion's Belt

Users access Orion's Belt via a graphical, browser-based interface. Currently, we support Internet Explorer on both PC and Mac platforms, and Netscape. User requests are

transmitted to a web server and thence to an application server, where database queries are converted to SQL and forwarded to an Oracle database engine.

Orion's Belt offers users three key features:

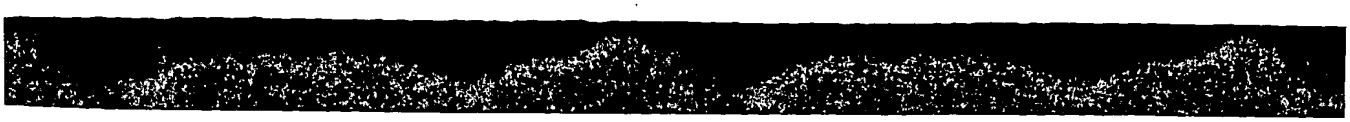
1. Connections
2. ClientLink
3. Browse

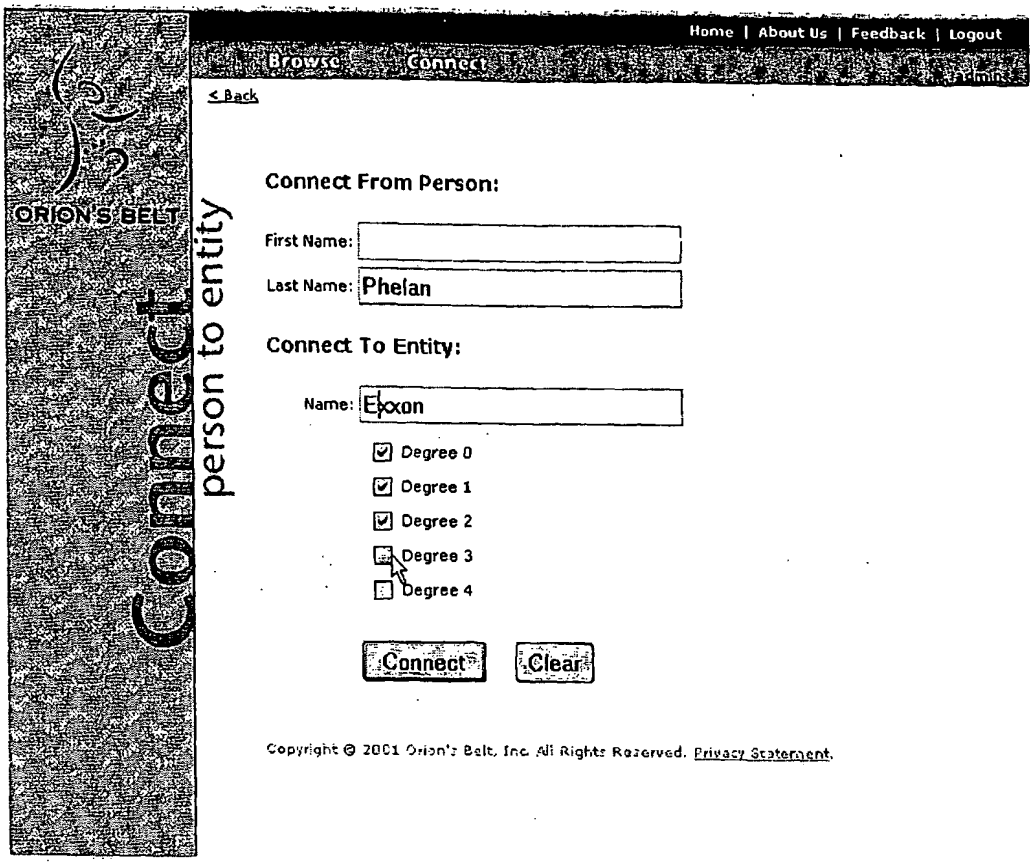
1. Connections

Users can ask the system to find connecting paths between a starting point (either a person or an entity) and an end point (which can also be either a person or an entity). Hence there are four connection possibilities:

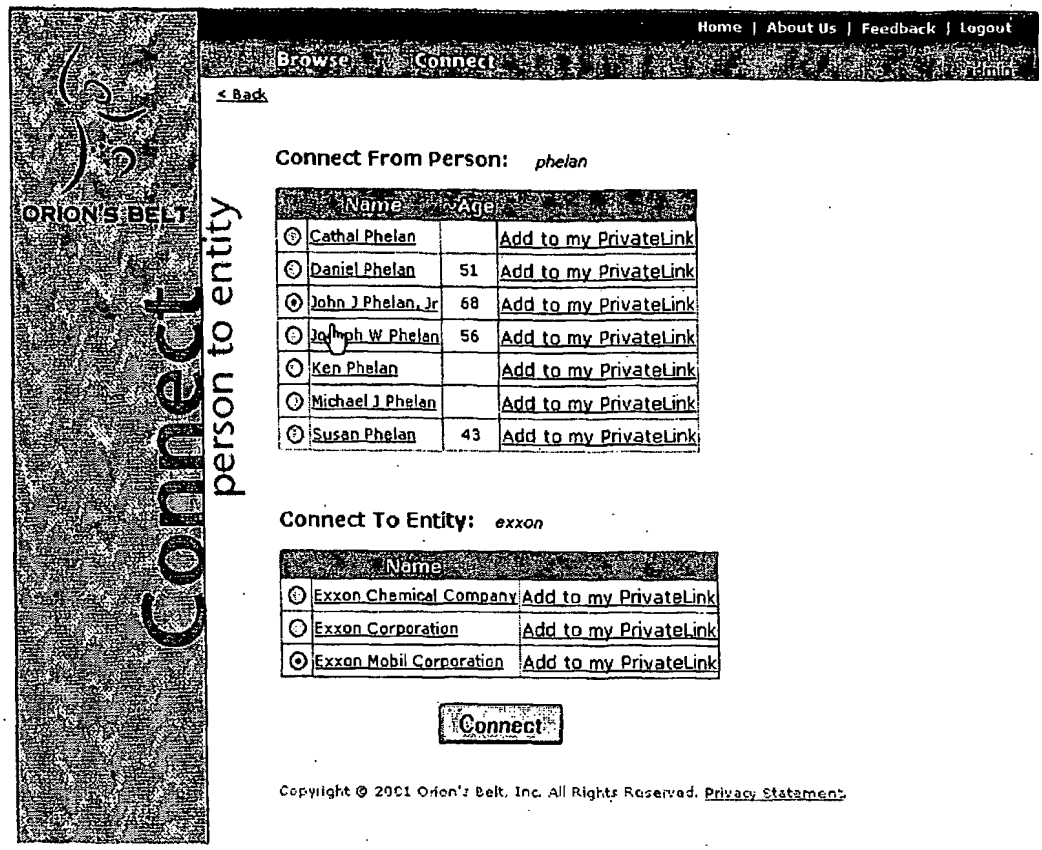
- Person to Person
- Person to Entity
- Entity to Person
- Entity to Entity

For example, suppose a user wanted to know if there was a path between John Phelan (a former chairman of the New York Stock Exchange) and Exxon Mobil Company. After requesting a Person-to-Entity connection, the user is asked to specify the person and the entity, as shown in the screen below:

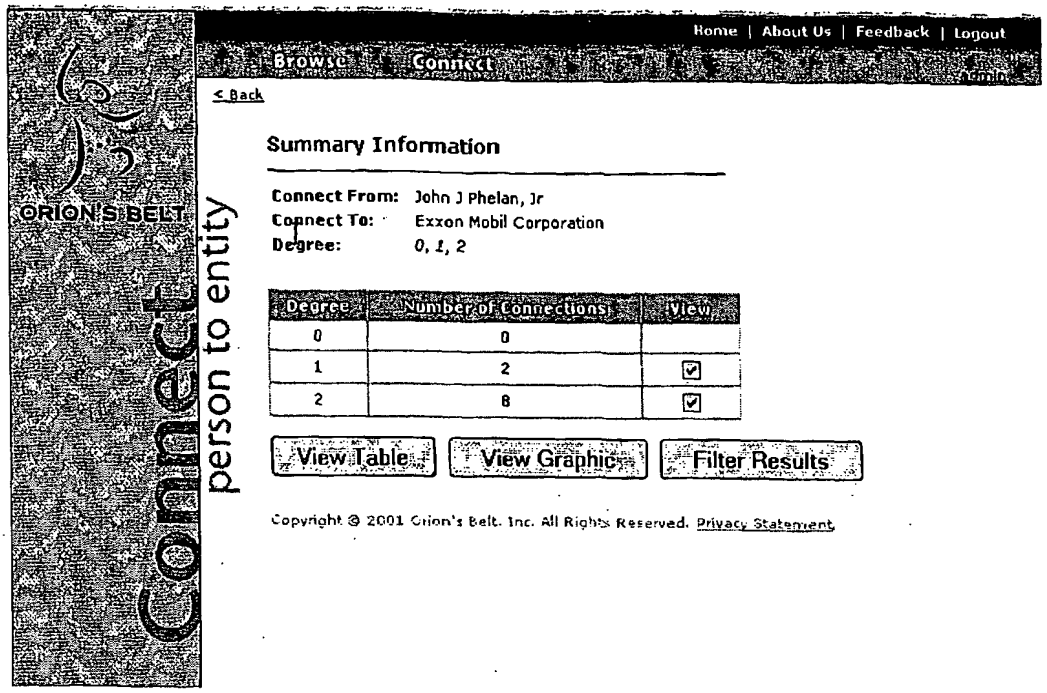




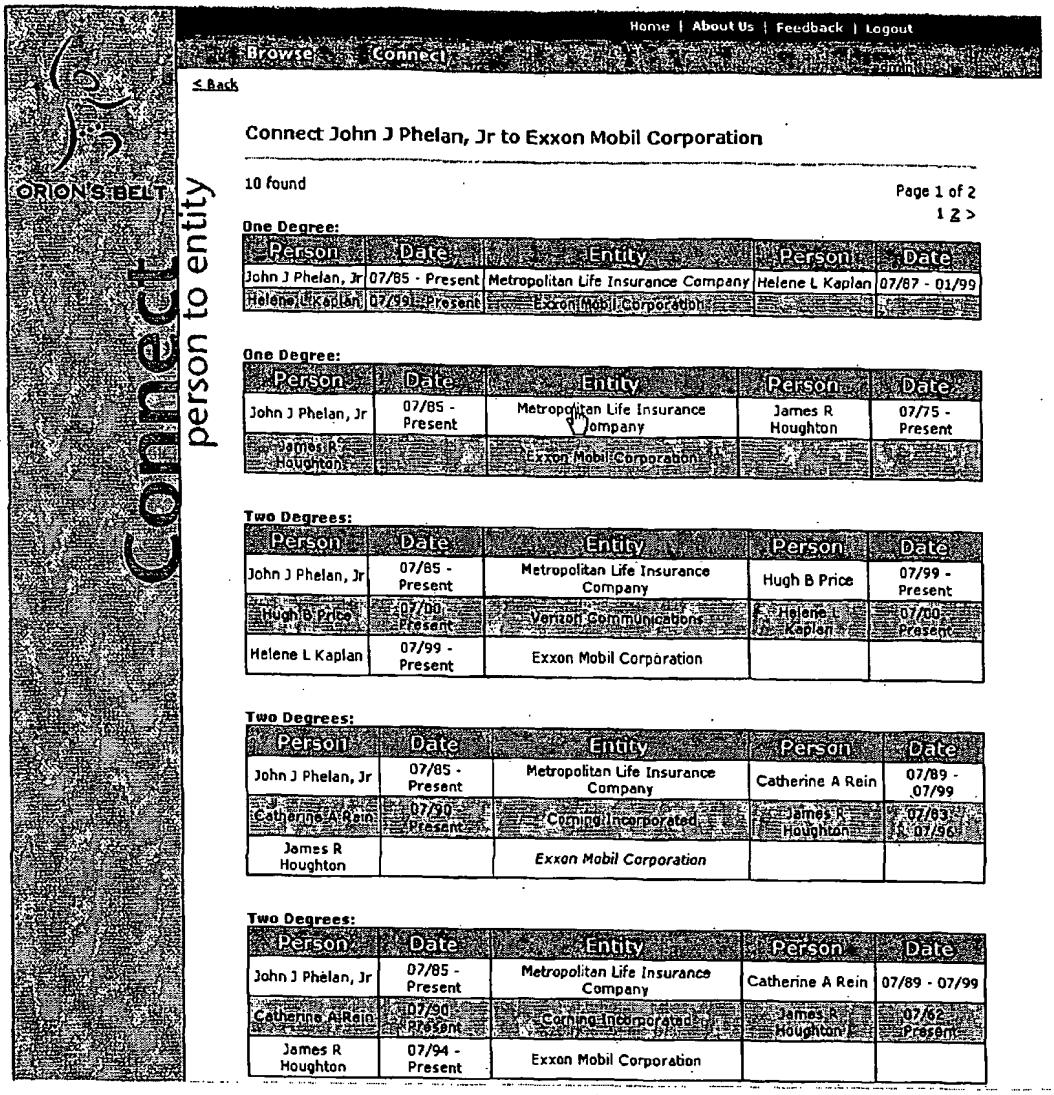
The degrees specify the number of people allowed between the starting and ending points. The next screen, shown below, allows the user to uniquely identify the desired person and entity from among several possible candidates:


ORION'S BELT
Connect
person to entity
Home | About Us | Feedback | Logout
Browse
Connect
< Back
Connect From Person: phelan
Name
Age
Cathal Phelan
Add to my PrivateLink
Daniel Phelan
51
Add to my PrivateLink
John J Phelan, Jr
68
Add to my PrivateLink
Joseph W Phelan
56
Add to my PrivateLink
Ken Phelan
Add to my PrivateLink
Michael J Phelan
Add to my PrivateLink
Susan Phelan
43
Add to my PrivateLink
Connect To Entity: exxon
Name
Exxon Chemical Company
Add to my PrivateLink
Exxon Corporation
Add to my PrivateLink
Exxon Mobil Corporation
Add to my PrivateLink
Connect
Copyright © 2001 Orion's Belt, Inc. All Rights Reserved. Privacy Statement

After selecting the particular person and entity desired, the system searches for possible paths between these starting and ending points, and displays a summary page, shown below:



Finally, the user can request one of three ways to view the results. The tabular view is shown below:



Facilities are also available for the user to filter the results, which is valuable when a large number of connections are returned. Going forward, we will also provide convenient ways to export connection information to desktop applications (e.g., MS Word, MS Excel, electronic mail).

2. ClientLink (see also attached FIG. 1 -11)

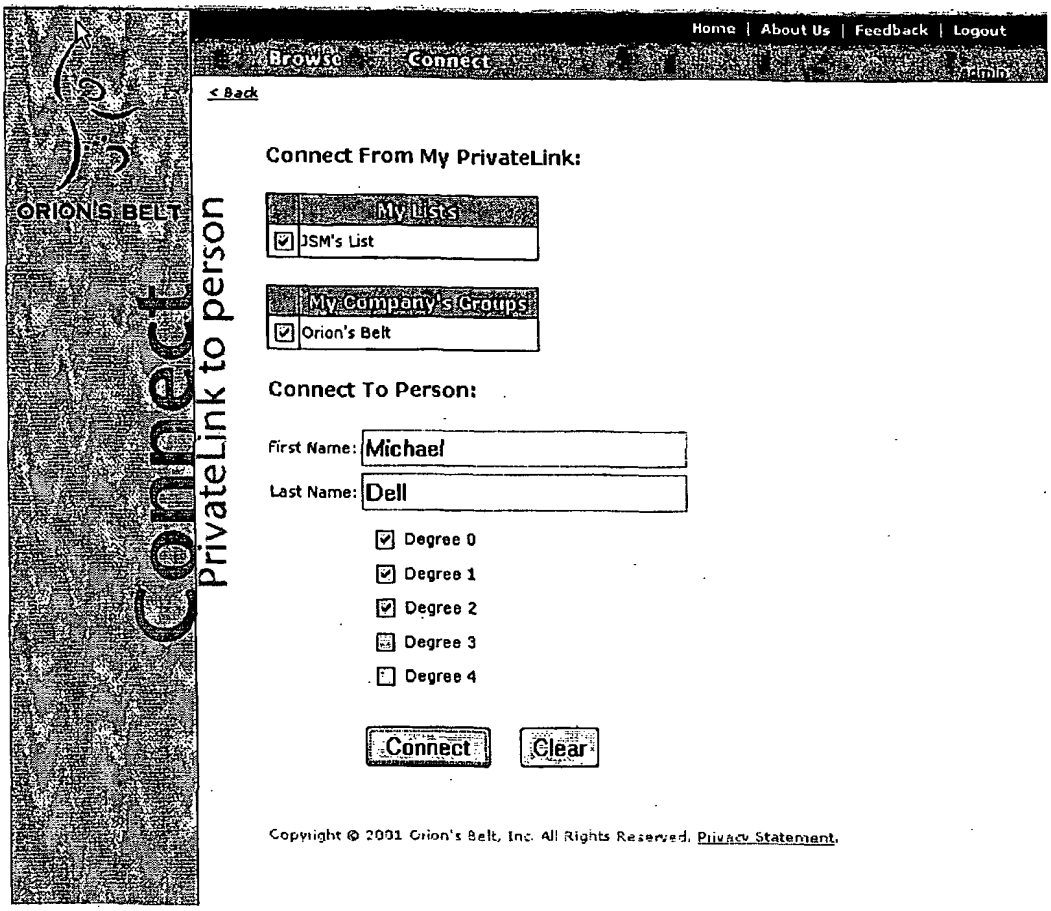
Client Link is a proprietary feature which links a client's own contacts (e.g., customers, vendors, etc.) with the OBI database in order to produce the most effective links for each

client. Essentially, this feature allows a client user to specify in advance the people or entities in the OBI database which are to be used as sources for a connection, thus eliminating the need to specify a unique starting point for each connection request.

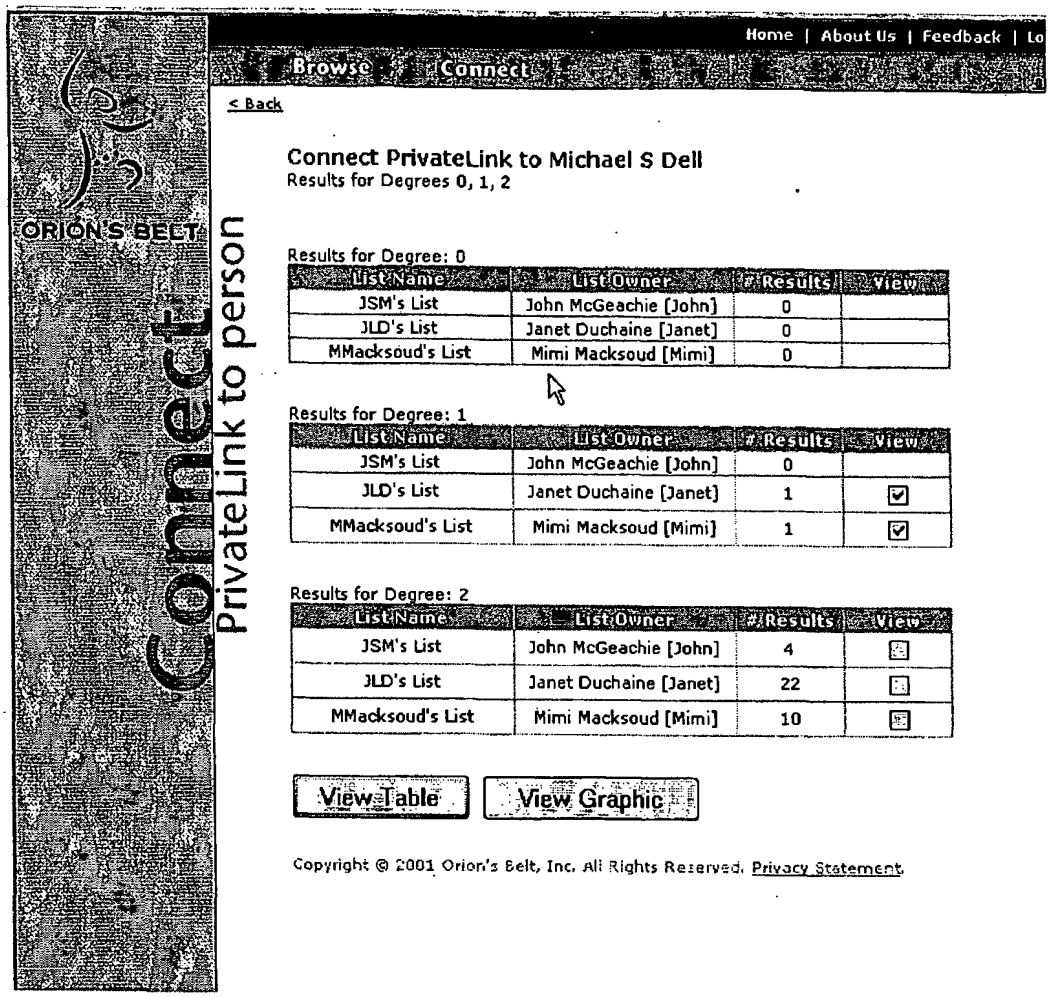
Individual users' lists can be part of a group, and connections can be requested using groups as a starting point. This feature allows client users to request connections from their own or from their colleagues' contacts. Client administrators have wide latitude in setting up groups, so that connections can be requested from an office, a region, a practice, or an entire organization. Security protocols prevent any client from accessing another client's ClientLink data.

OBI will offer consulting services to customize ClientLink for each client during the installation process. These services will include, for example, determining the most effective way to make existing contact lists (e.g., from common contact list or CRM products) accessible to the OBI system.

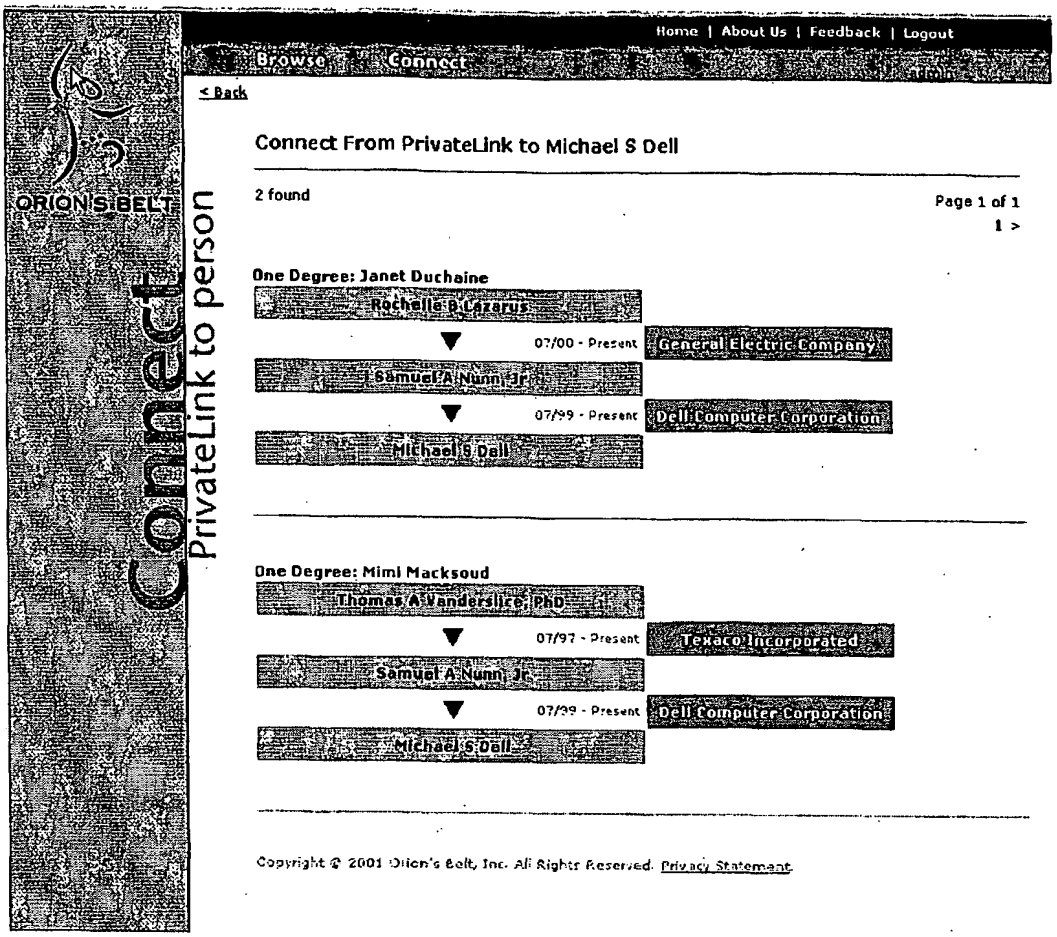
Within the system interface, a user's ClientLink list is called a PrivateLink. Users can request connections from their PrivateLink to either a person or an entity. An example is shown below, requesting a connection from our (Orion's Belt) Private Link to Michael Dell, of Dell Computer Corporation:



The summary display (below) shows the "owners" of the lists that have connections. User can avail themselves of both their own and their colleagues' connections to maximize the likelihood of finding a reasonable link to a designated destination. A user's ability to access Private Link lists other than his or her own is determined by the client administrator. This particular example shows how each of us at Orion's Belt, Inc., can make use of our colleagues' lists to find the best path to the desired end-point.



The final result screen displays the connections from within each list, shown below using the graphical view:



Users can populate their Private Links by using our services to extract data from their current contact lists, or they can easily enter data into their Private Links as they use our system. We expect to offer add-on modules to enable the Orion's Belt service to access users' existing contact lists directly.

4. Browse

The browse function enables a user either to look at people, and the entities with which they are associated, or to look at entities, and the people associated with them. The browse function can be invoked at any time during the connection process; there is also a separate browse function on the main menu.

The browse function can be extended to include an "explore" function, which begins with the endpoint (person or entity) and display progressively larger circles of contacts, so that the user might look for known contacts.

Connection Technology Extensions

The connection algorithms look for overlaps between the time periods during which two or more people were associated with an entity. But the connection algorithms themselves have no intrinsic knowledge of people and entities—they actually look for overlaps between entries in a general-purpose relational database. These entries could be, for example:

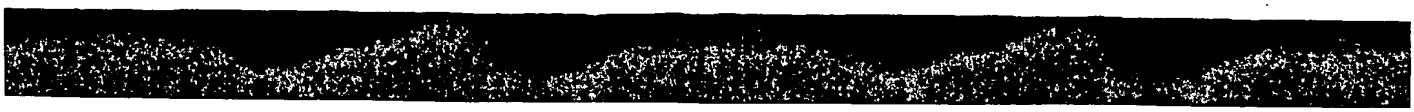
- Web sites and their visitors
- Trucks and their cargo
- Airline flights and their passengers

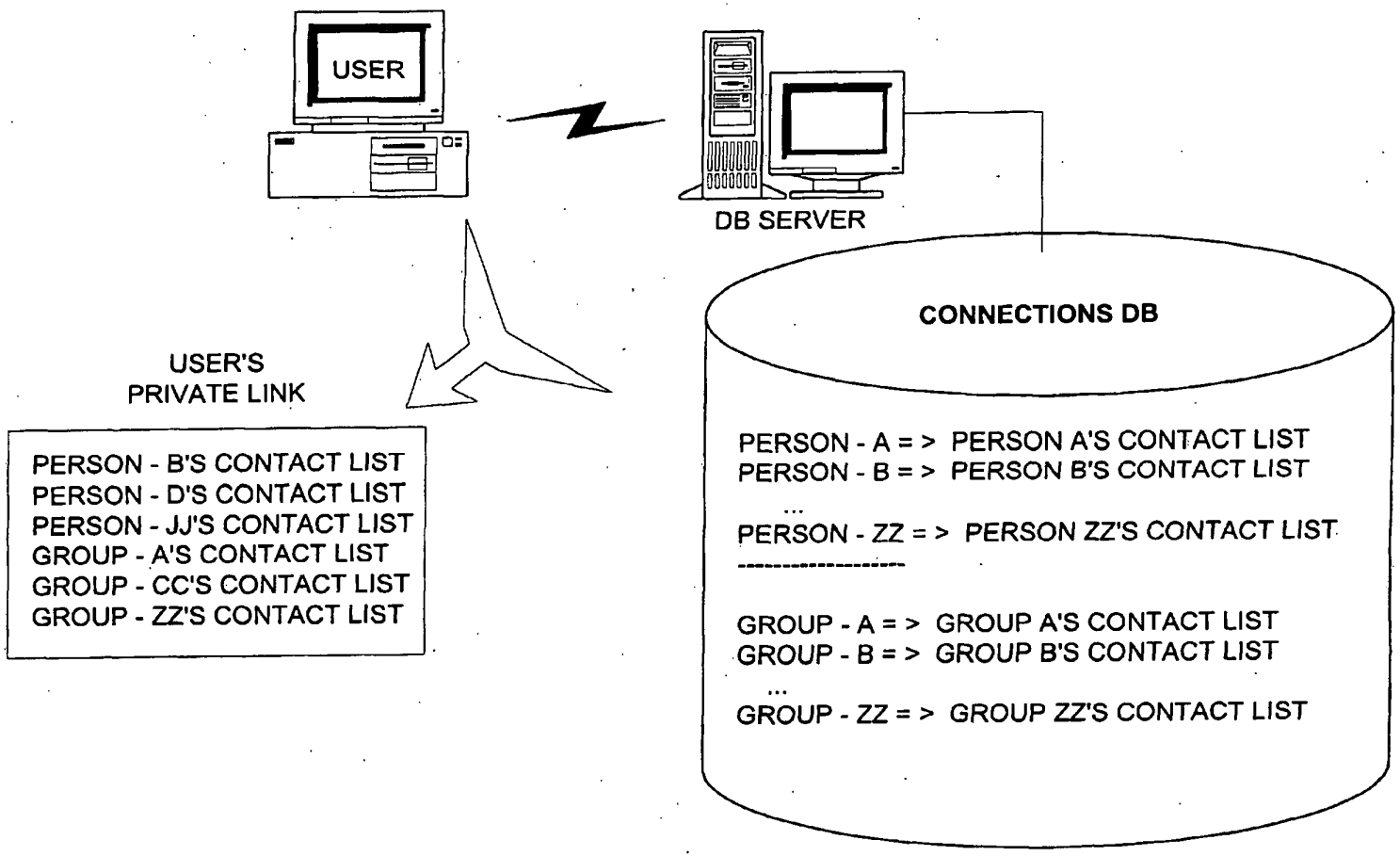
More generally, entries in the database can represent containers or contents-of-containers, where a content entry is associated with a container entry over some (perhaps indefinite) period of time. Containers can themselves be the contents of other containers.

The connection technology and associated user interface can also be applied to clients' private databases (e.g., a recruiting firm's inventory of potential candidates). Third-party databases can be integrated into the Orion's Belt service, permitting revenue sharing arrangements with established content providers.

Parsing Technology

As described earlier, the parser is uniquely able to identify entities, people, and the relationship between entity-people pairs. There are several examples of seemingly similar technologies, such as lexical analyzers which deconstruct sentences into their components, or search engines which index all or selected words from web pages. However, none of these technologies that we have evaluated to-date is able to match up people with entities.

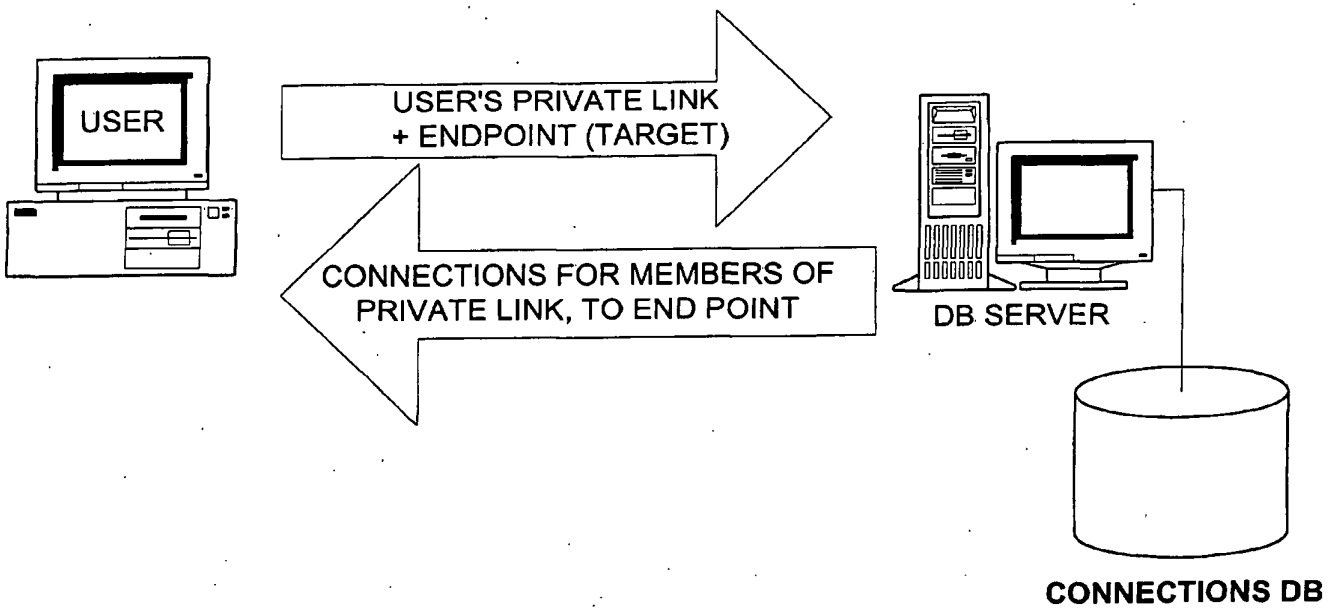




PRIVATE LINK

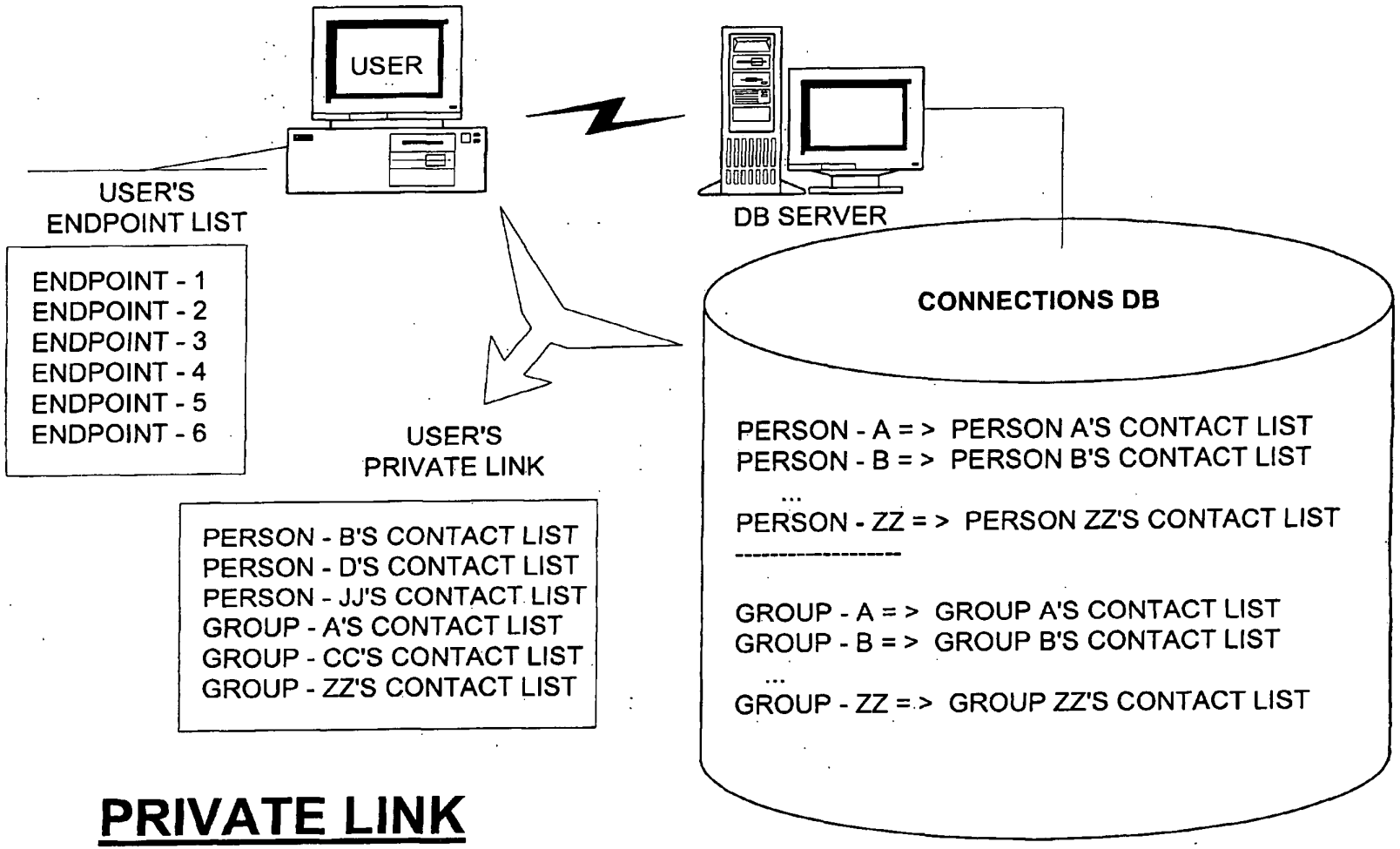
NOTE: A USER'S PRIVATELINK LIST CAN BE RESIDENT ON THE CONNECTIONS SYSTEM, USER SYSTEM, THIRD PARTY SYSTEM, OR SOME COMBINATION THEREOF.

FIGURE 1



PRIVATE LINK: QUERY & RESULTS

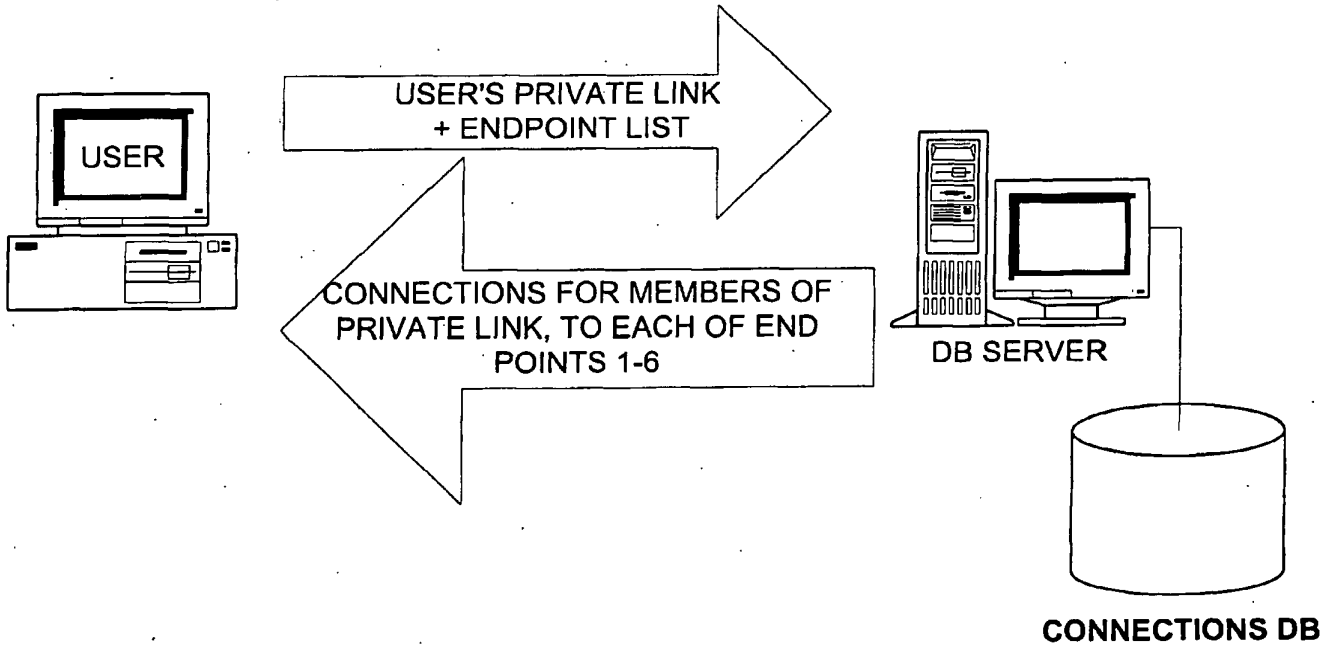
FIGURE 2



PRIVATE LINK WITH ENDPOINT LIST

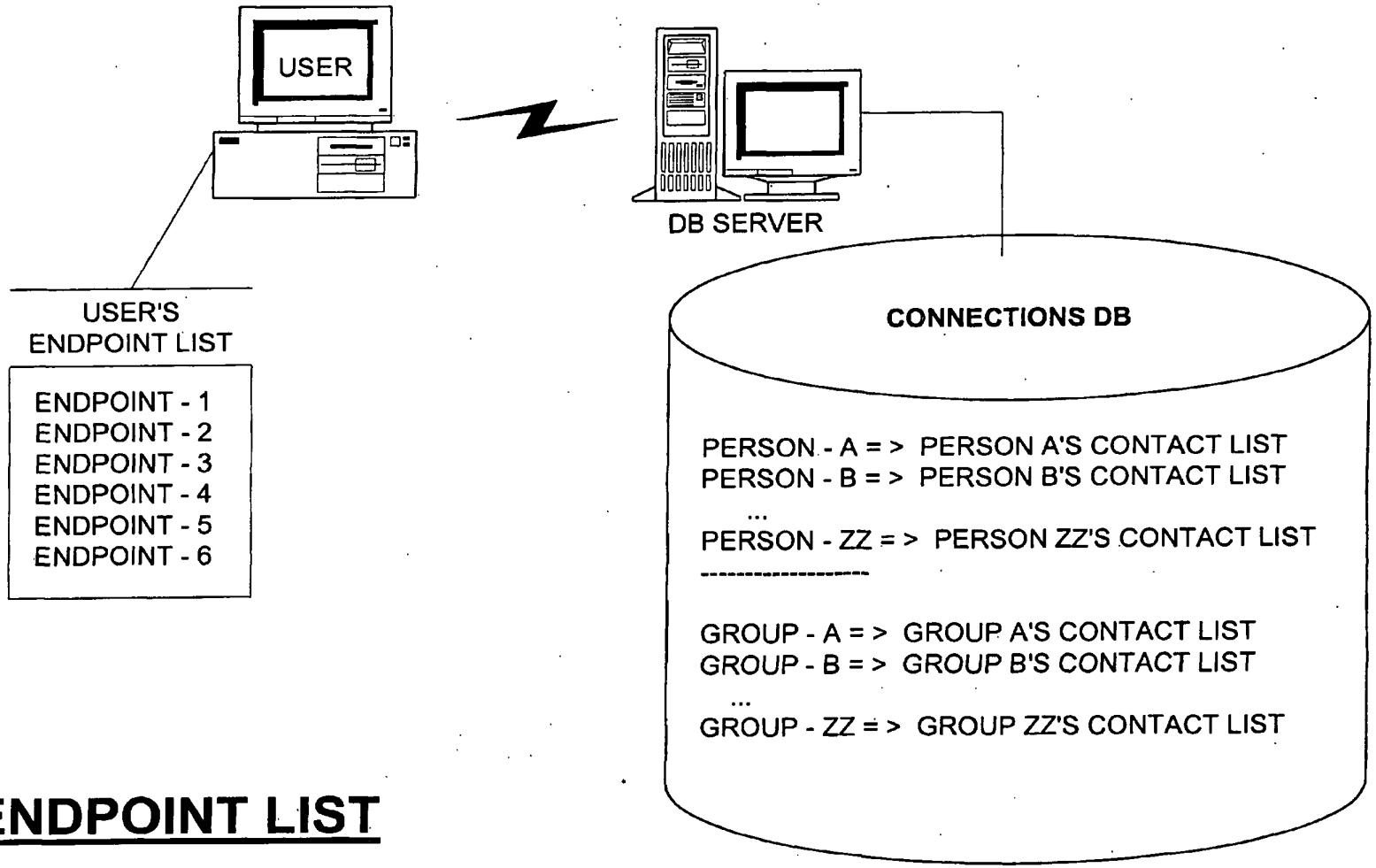
NOTE: A USER'S PRIVATELINK LIST & ENDPOINT LIST CAN BE RESIDENT ON THE CONNECTIONS SYSTEM, USER SYSTEM, THIRD PARTY SYSTEM, OR SOME COMBINATION THEREOF.

FIGURE 3



PRIVATE LINK WITH ENDLIST: QUERY & RESULTS

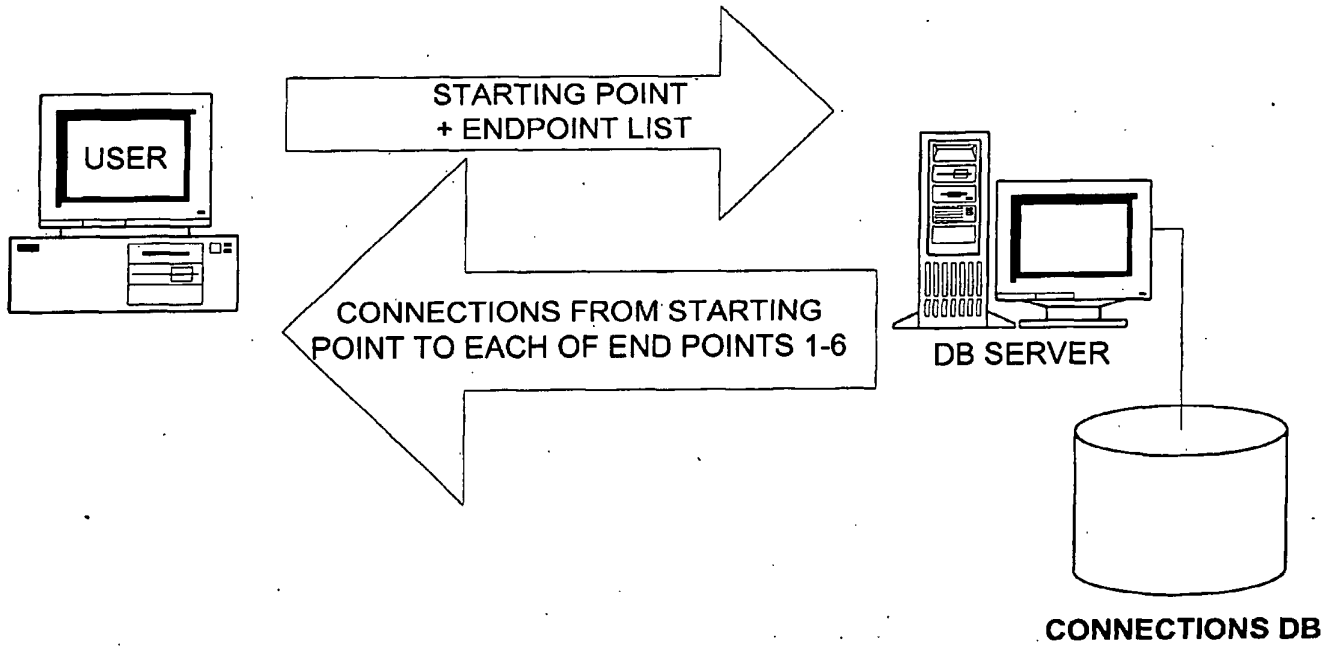
FIGURE 4



ENDPOINT LIST

NOTE: A USER'S ENDPOINT LIST CAN BE RESIDENT ON THE CONNECTIONS SYSTEM, USER SYSTEM, THIRD PARTY SYSTEM, OR SOME COMBINATION THEREOF.

FIGURE 5



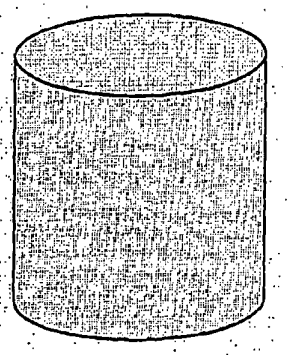
ENDPOINT LIST: QUERY & RESULTS

FIGURE 6

To Create PrivateLink...

Orion's Belt Database

Client's Database
or CRM System



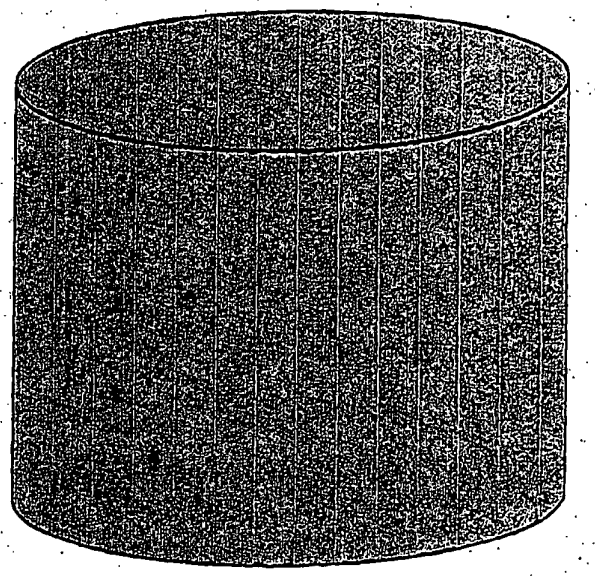
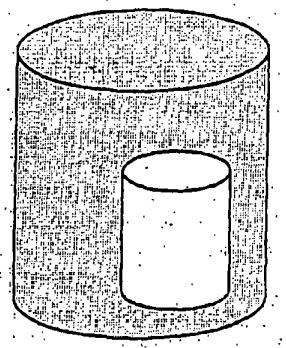


FIG 7

To Create PrivateLink...

Orion's Belt Database

Client's Database
or CRM System



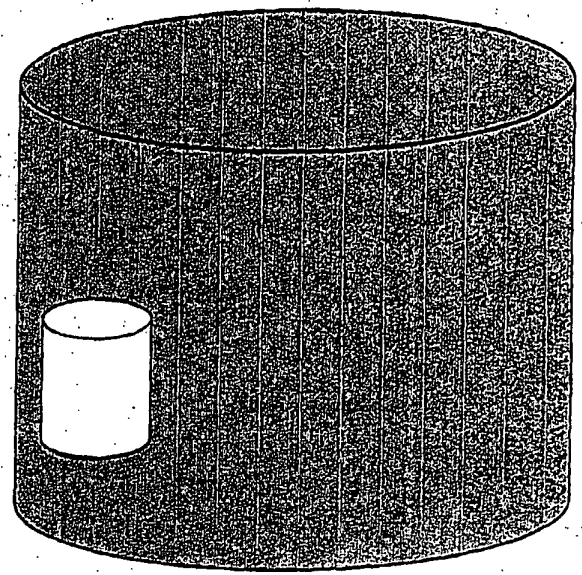
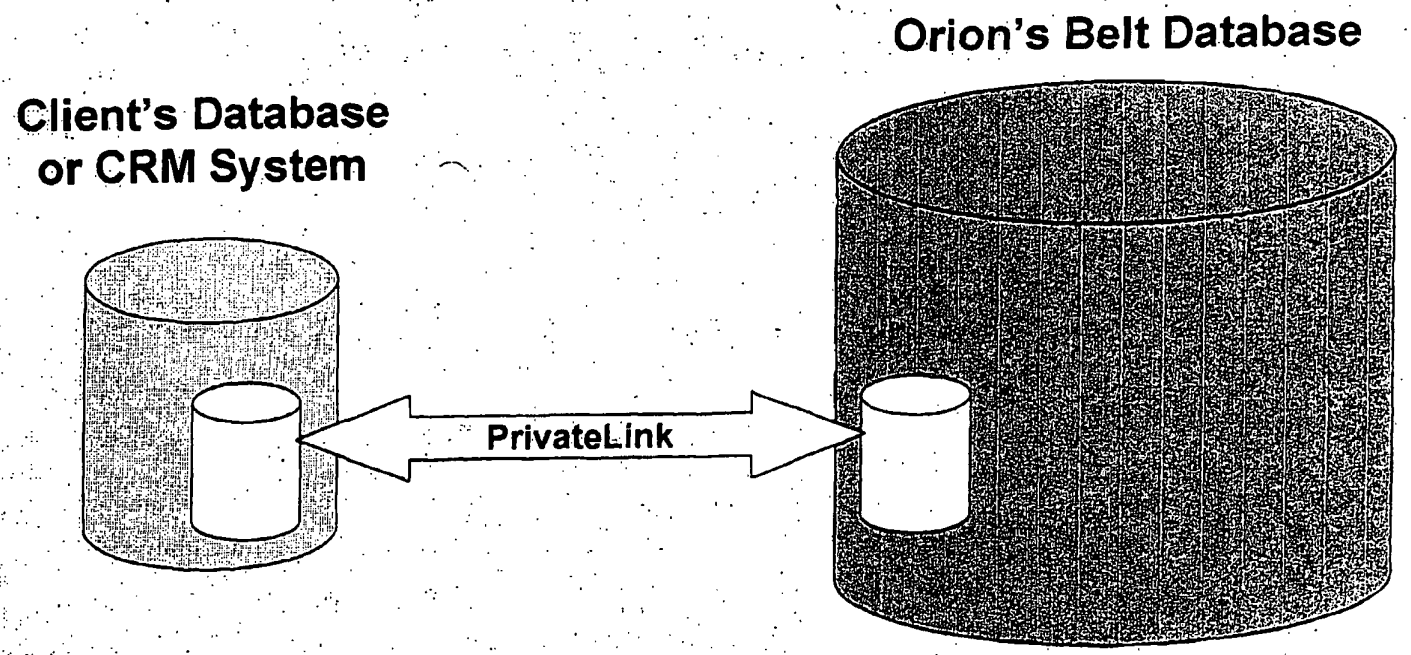


Fig 8

To Create PrivateLink...



EG-9

To Create PrivateLink...

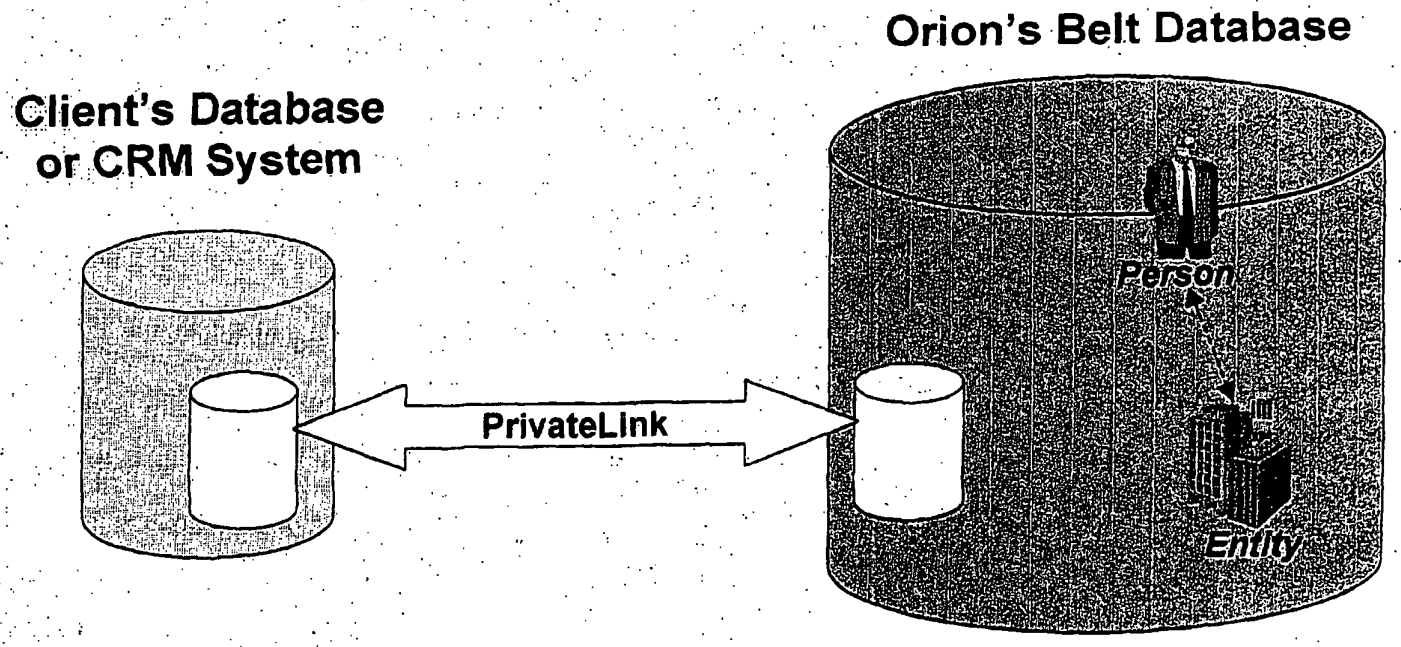


FIG. 10

To Create PrivateLink...

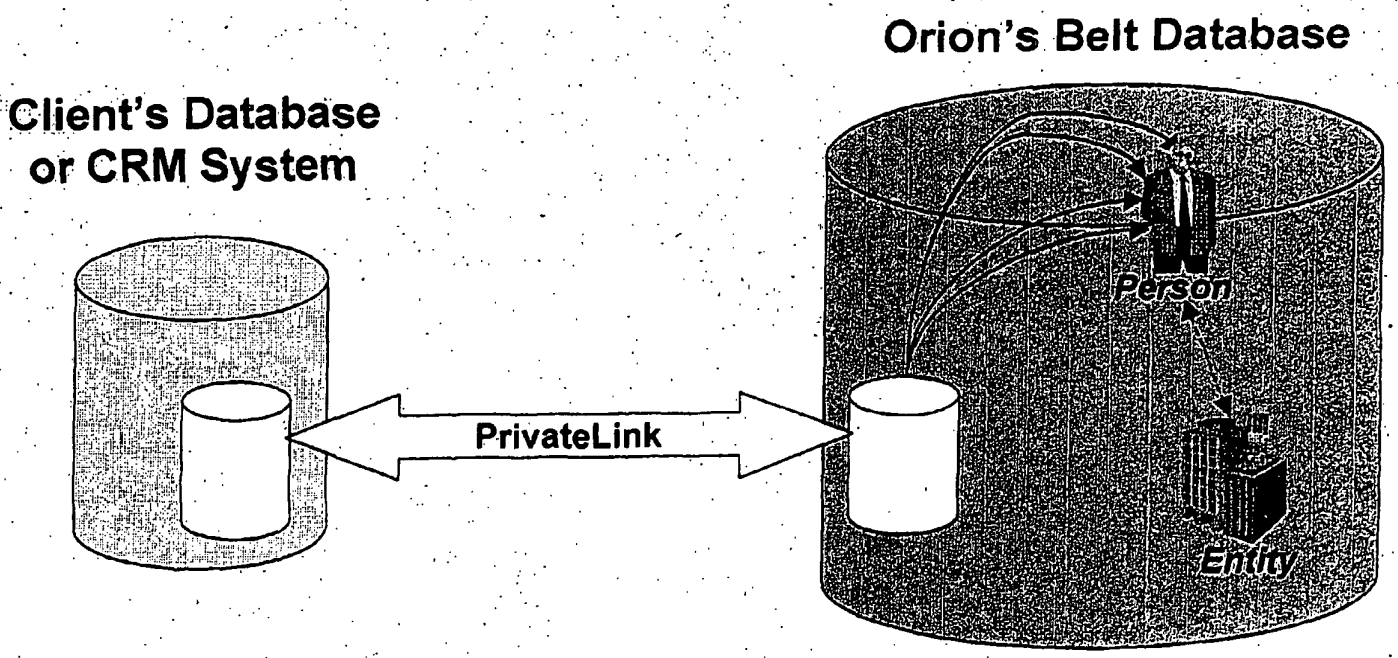
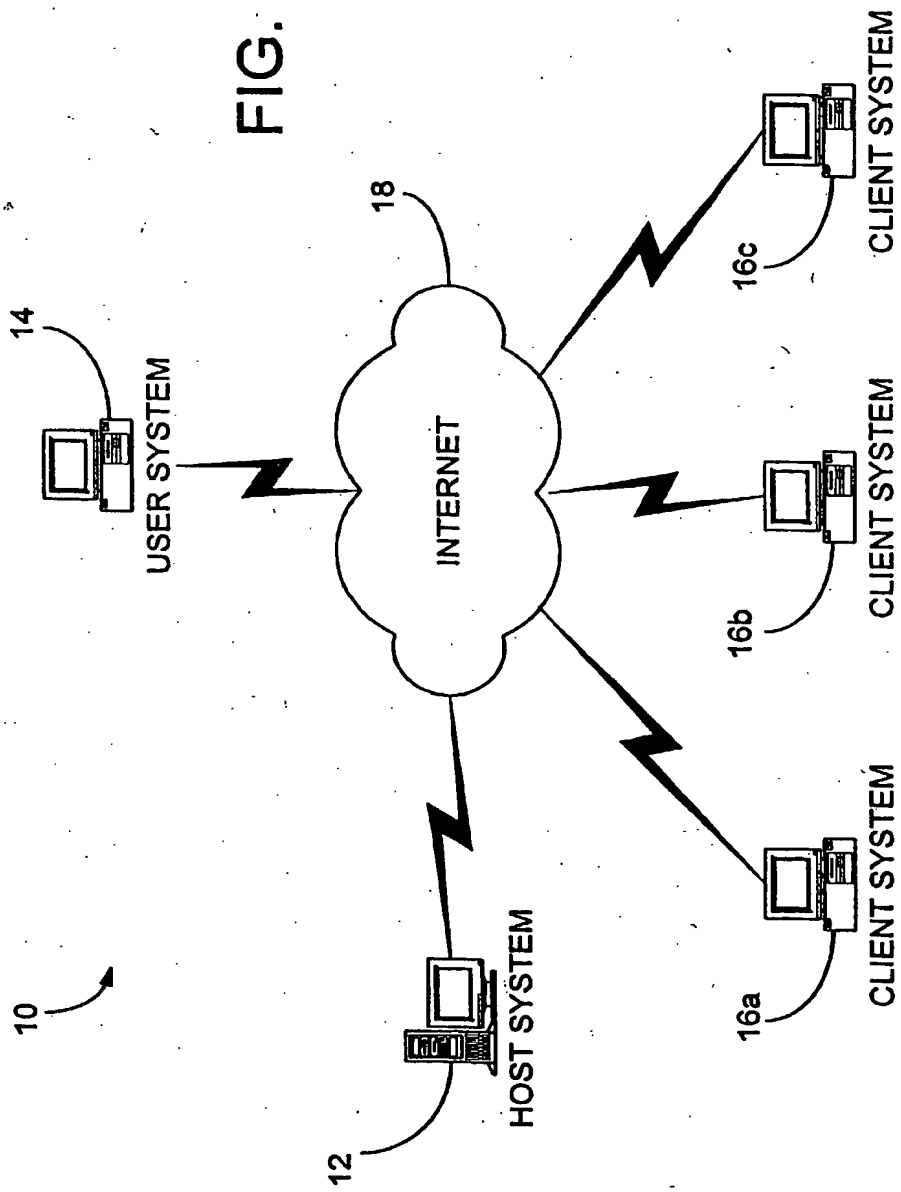


FIG. 1

FIG. 1


10
12
HOST SYSTEM
14
USER SYSTEM
18
INTERNET
16a
CLIENT SYSTEM
16b
CLIENT SYSTEM
16c
CLIENT SYSTEM

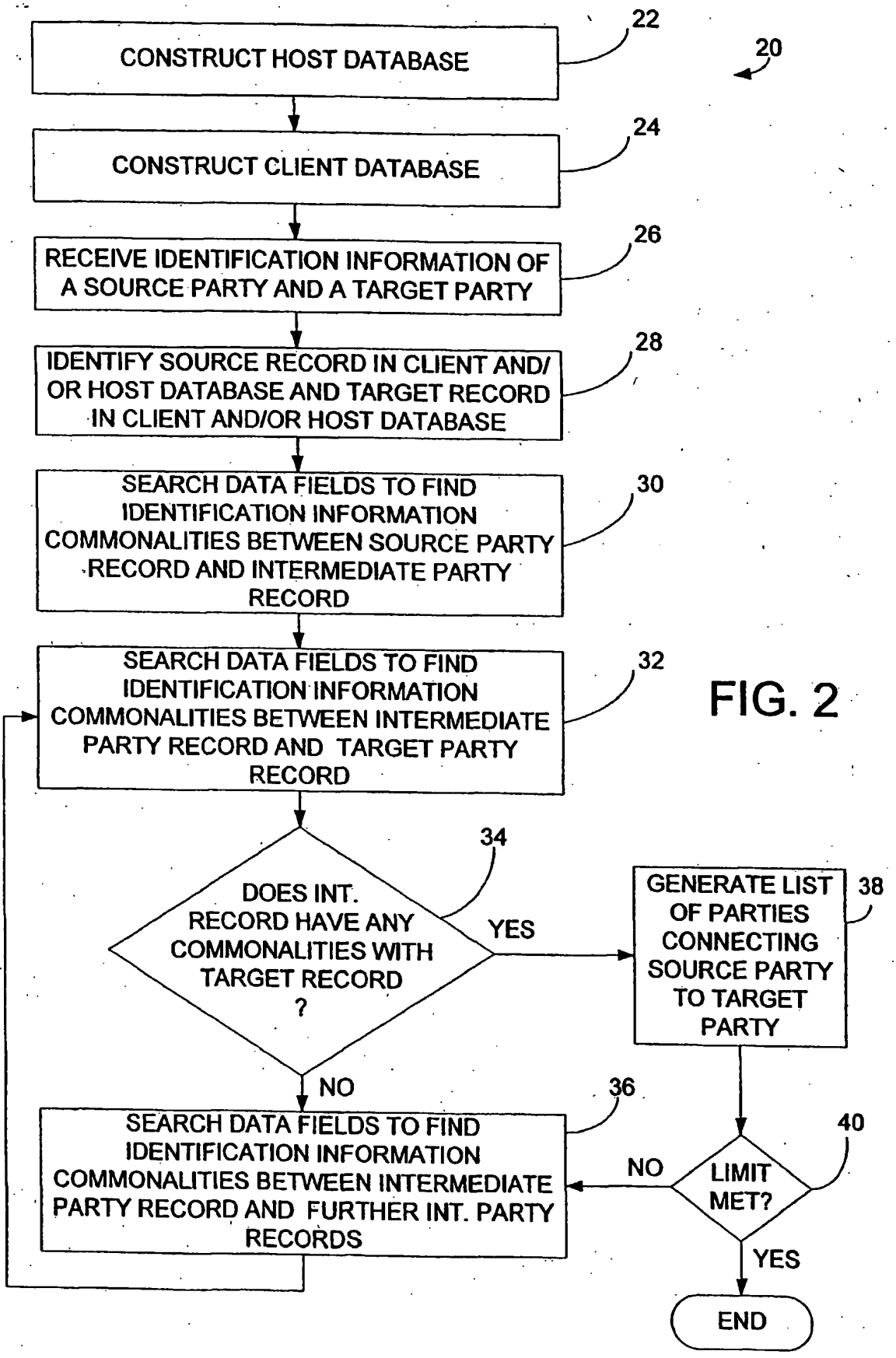

20
CONSTRUCT HOST DATABASE
22
CONSTRUCT CLIENT DATABASE
24
RECEIVE IDENTIFICATION INFORMATION OF A SOURCE PARTY AND A TARGET PARTY
26
IDENTIFY SOURCE RECORD IN CLIENT AND/ OR HOST DATABASE AND TARGET RECORD IN CLIENT AND/OR HOST DATABASE
28
SEARCH DATA FIELDS TO FIND IDENTIFICATION INFORMATION COMMONALITIES BETWEEN SOURCE PARTY RECORD AND INTERMEDIATE PARTY RECORD
30
SEARCH DATA FIELDS TO FIND IDENTIFICATION INFORMATION COMMONALITIES BETWEEN INTERMEDIATE PARTY RECORD AND TARGET PARTY RECORD
32
DOES INT. RECORD HAVE ANY COMMONALITIES WITH TARGET RECORD ?
34
YES
GENERATE LIST OF PARTIES CONNECTING SOURCE PARTY TO TARGET PARTY
38
NO
SEARCH DATA FIELDS TO FIND IDENTIFICATION INFORMATION COMMONALITIES BETWEEN INTERMEDIATE PARTY RECORD AND FURTHER INT. PARTY RECORDS
36
LIMIT MET?
40
NO
YES
END

FIG. 2

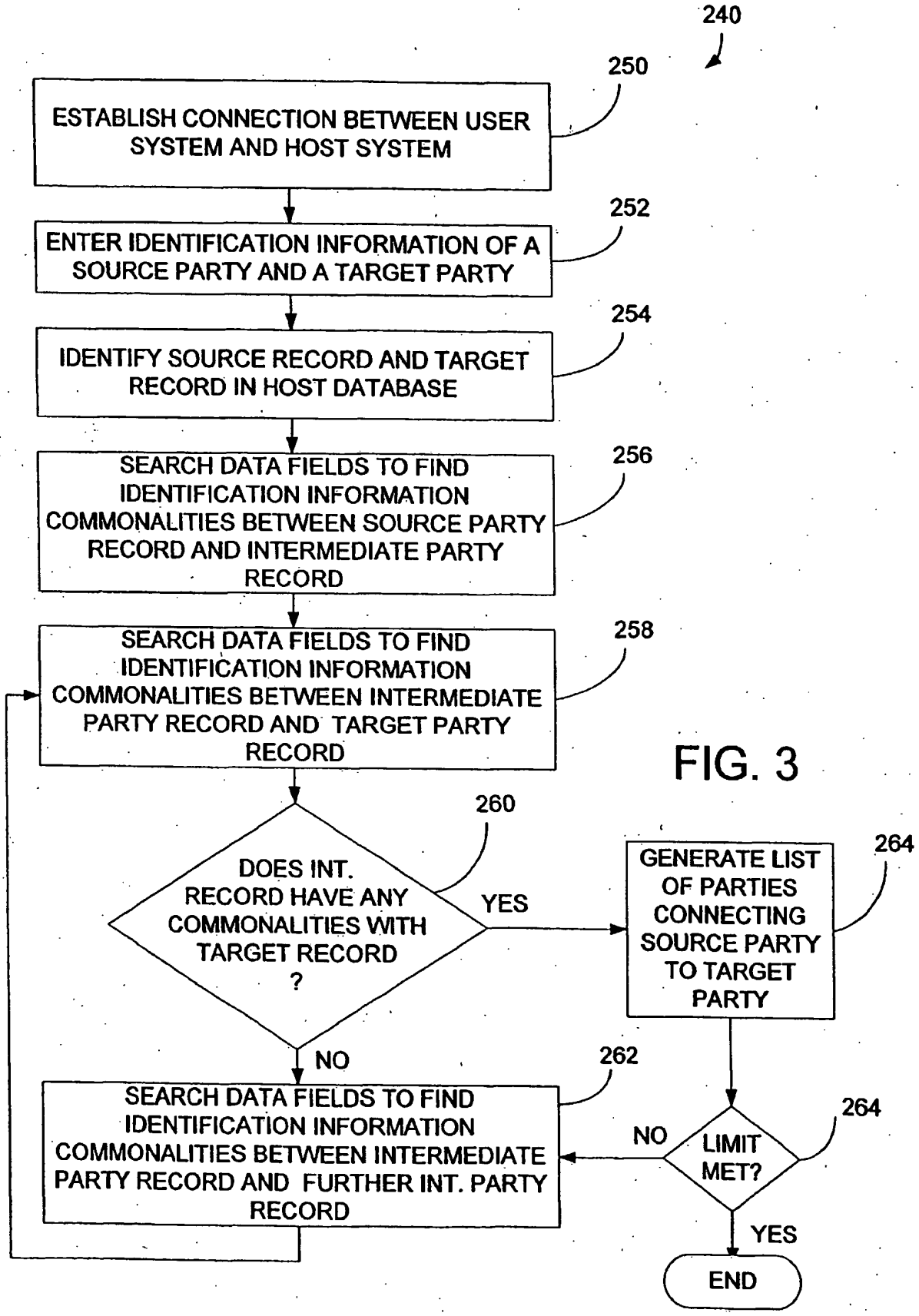
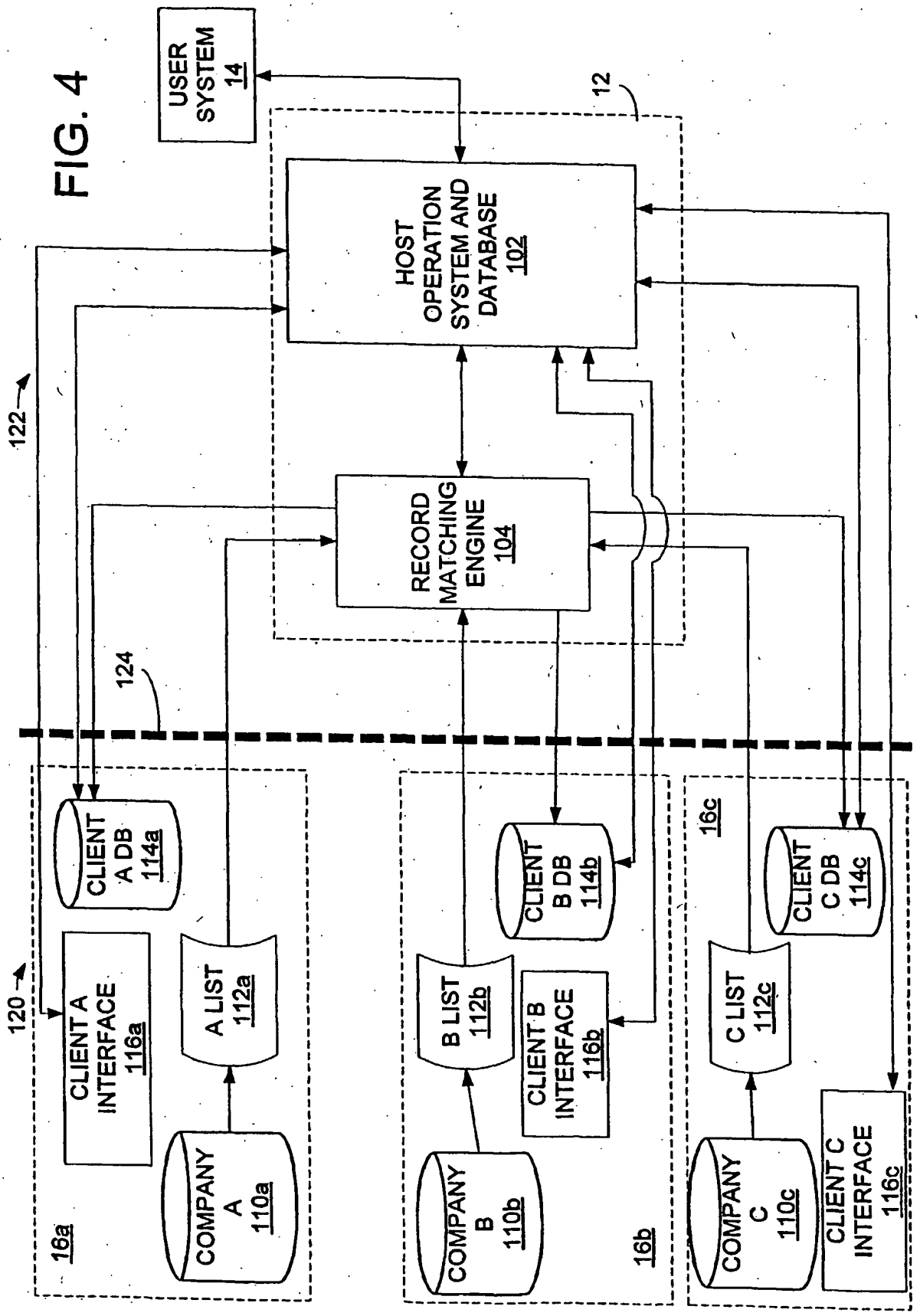

240
250
ESTABLISH CONNECTION BETWEEN USER SYSTEM AND HOST SYSTEM
252
ENTER IDENTIFICATION INFORMATION OF A SOURCE PARTY AND A TARGET PARTY
254
IDENTIFY SOURCE RECORD AND TARGET RECORD IN HOST DATABASE
256
SEARCH DATA FIELDS TO FIND IDENTIFICATION INFORMATION COMMONALITIES BETWEEN SOURCE PARTY RECORD AND INTERMEDIATE PARTY RECORD
258
SEARCH DATA FIELDS TO FIND IDENTIFICATION INFORMATION COMMONALITIES BETWEEN INTERMEDIATE PARTY RECORD AND TARGET PARTY RECORD
260
DOES INT. RECORD HAVE ANY COMMONALITIES WITH TARGET RECORD ?
YES
NO
264
GENERATE LIST OF PARTIES CONNECTING SOURCE PARTY TO TARGET PARTY
262
SEARCH DATA FIELDS TO FIND IDENTIFICATION INFORMATION COMMONALITIES BETWEEN INTERMEDIATE PARTY RECORD AND FURTHER INT. PARTY RECORD
264
LIMIT MET?
NO
YES
END

FIG. 3

FIG. 4


USER SYSTEM 14
12
HOST OPERATION SYSTEM AND DATABASE 102
122
RECORD MATCHING ENGINE 104
124
120
CLIENT A INTERFACE 116a
A LIST 112a
CLIENT A DB 114a
COMPANY A 110a
16a
B LIST 112b
CLIENT B INTERFACE 116b
CLIENT B DB 114b
COMPANY B 110b
16b
16c
C LIST 112c
CLIENT C DB 114c
COMPANY C 110c
CLIENT C INTERFACE 116c

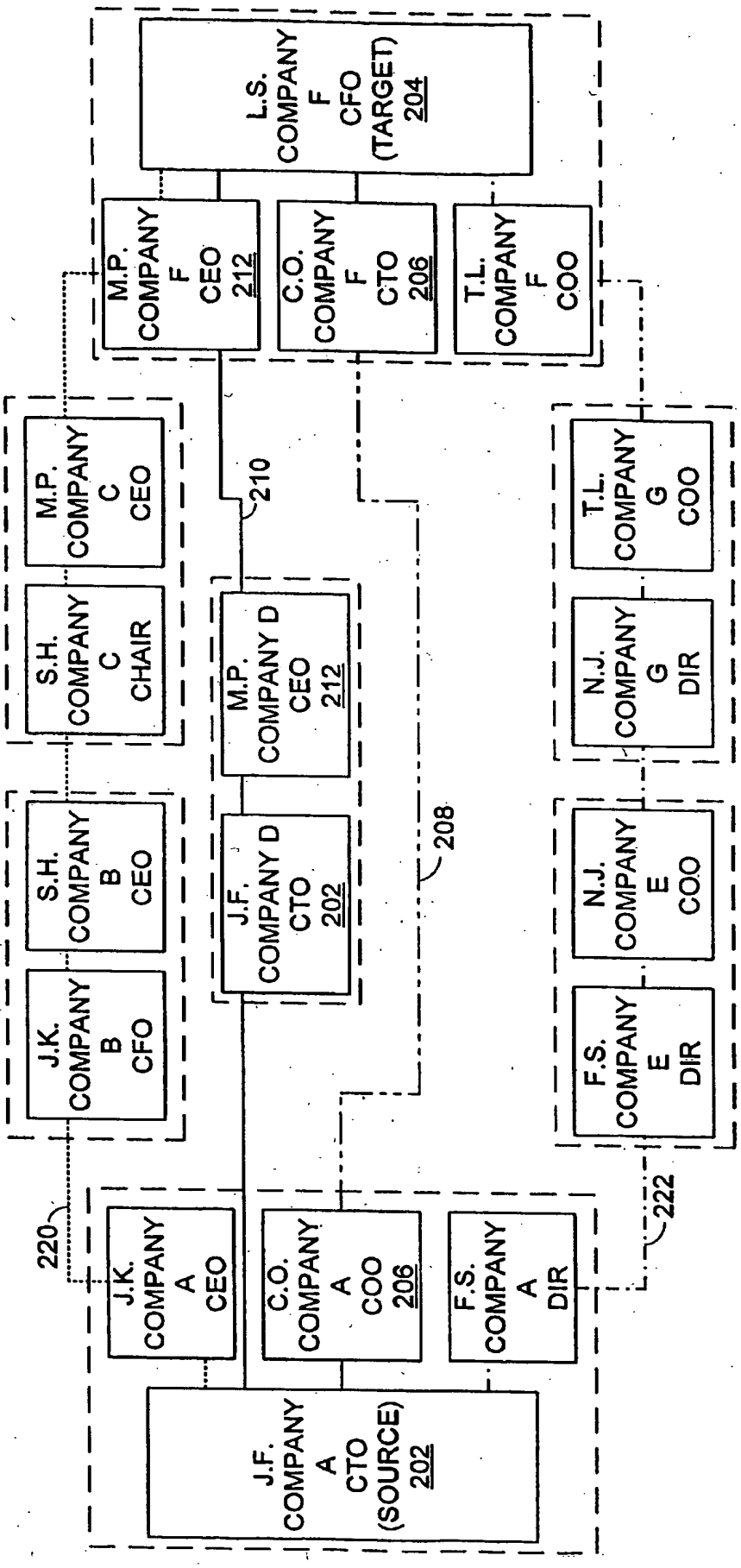

L.S. COMPANY F CFO (TARGET) 204
M.P. COMPANY F CEO 212
C.O. COMPANY F CTO 206
T.L. COMPANY F COO
M.P. COMPANY C CEO
S.H. COMPANY C CHAIR
M.P. COMPANY D CEO 212
T.L. COMPANY G COO
N.J. COMPANY G DIR
S.H. COMPANY B CEO
J.F. COMPANY D CTO 202
N.J. COMPANY E COO
J.K. COMPANY B CFO
F.S. COMPANY E DIR
210
208
220
222
J.K. COMPANY A CEO
C.O. COMPANY A COO 206
F.S. COMPANY A DIR
J.F. COMPANY A CTO (SOURCE) 202

FIG. 5

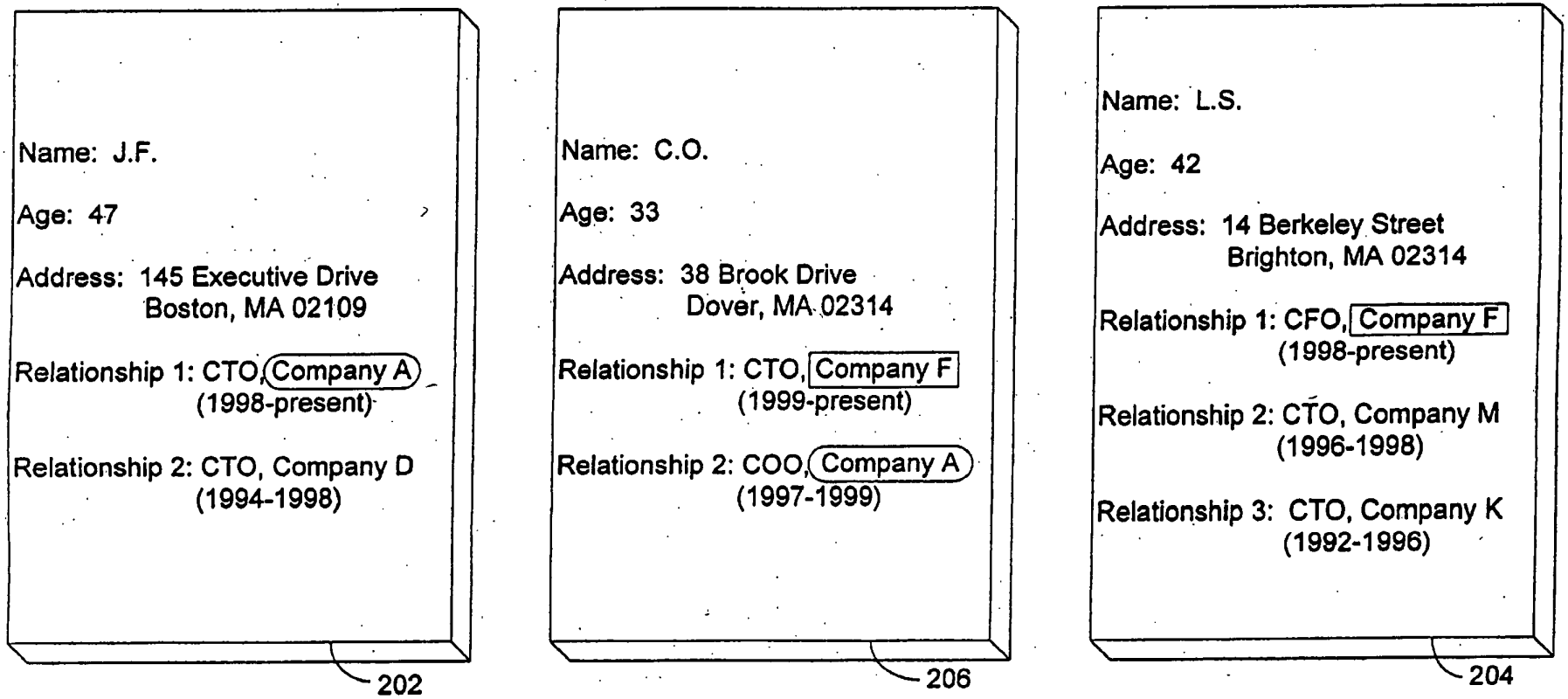

Name: J.F.
Age: 47
Address: 145 Executive Drive
Boston, MA 02109
Relationship 1: CTO, Company A
(1998-present)
Relationship 2: CTO, Company D
(1994-1998)
202
Name: C.O.
Age: 33
Address: 38 Brook Drive
Dover, MA 02314
Relationship 1: CTO, Company F
(1999-present)
Relationship 2: COO, Company A
(1997-1999)
206
Name: L.S.
Age: 42
Address: 14 Berkeley Street
Brighton, MA 02314
Relationship 1: CFO, Company F
(1998-present)
Relationship 2: CTO, Company M
(1996-1998)
Relationship 3: CTO, Company K
(1992-1996)
204

FIG. 6

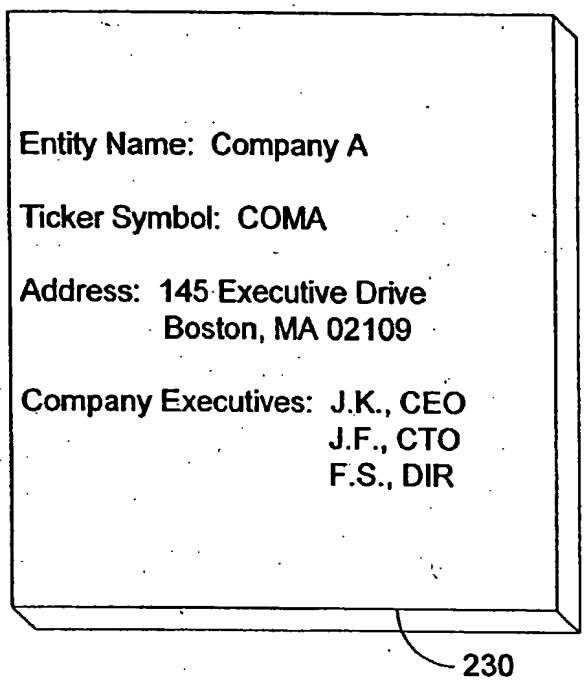

Entity Name: Company A
Ticker Symbol: COMA
Address: 145 Executive Drive
Boston, MA 02109
Company Executives: J.K., CEO
J.F., CTO
F.S., DIR
230

FIG. 7